

This electronic thesis or dissertation has been downloaded from the King's Research Portal at <https://kclpure.kcl.ac.uk/portal/>



Call Admission Control in Wireless Mesh Networks

Naghavi, Nika

Awarding institution:
King's College London

The copyright of this thesis rests with the author and no quotation from it or information derived from it may be published without proper acknowledgement.

END USER LICENCE AGREEMENT



Unless another licence is stated on the immediately following page this work is licensed

under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International

licence. <https://creativecommons.org/licenses/by-nc-nd/4.0/>

You are free to copy, distribute and transmit the work

Under the following conditions:

- Attribution: You must attribute the work in the manner specified by the author (but not in any way that suggests that they endorse you or your use of the work).
- Non Commercial: You may not use this work for commercial purposes.
- No Derivative Works - You may not alter, transform, or build upon this work.

Any of these conditions can be waived if you receive permission from the author. Your fair dealings and other rights are in no way affected by the above.

Take down policy

If you believe that this document breaches copyright please contact librarypure@kcl.ac.uk providing details, and we will remove access to the work immediately and investigate your claim.

Call Admission Control in Wireless Mesh Networks



Nika Naghavi

Centre for Telecommunications Research
School of Natural and Mathematical Sciences
King's College London

A thesis submitted to King's College London
in partial fulfillment of the requirements
for the degree of Doctor of Philosophy
January 2017

Abstract

Efficiency in managing scarce wireless resources has been a major design obstacle in wireless mobile networks since its advent. A large body of work has been published on the subject in relation to the first two generation and third generation (3G) of wireless networks. However, these solutions are not readily applicable to the future IP based multi-hop wireless networks. Increased trend in the number of users accessing multimedia rich traffic, and the impacts of mobility management support mechanisms alongside the shared nature of backhaul nodes demand exploration of new avenues of resource management policies. In this thesis a set of schemes are proposed to address the issue of Call Admission Control (CAC) in future multi-hop IP based mobile networks. The first scheme is based on a joint CAC and route assignment design mechanism in Wireless Mesh Networks (WMNs) environment, addressing issues such as pricing policies, wireless resource constraints of the access and backhaul links as well as maximising network provider's revenue. Study's focus is then shifted towards CAC design in Mobility Agent (MA) based Proxy Mobile IPv6 (PMIPv6) networks. A novel class-based CAC mechanism was proposed with the aim of eliminating bottleneck effect at the MAs. To further improve the overall capacity of the network, in conjunction with the proposed CAC mechanism, a modified version of a pre-

viously proposed Route Optimisation (RO) scheme is implemented. The performance of the proposed schemes is investigated and analysed through extensive simulations. Comparative studies are then carried out to examine the network provider's total revenue gain in WMNs environment, total blocking probability and per class blocking probability in PMIPv6 networks.

Acknowledgements

This thesis summarises my time as a PhD candidate. It has been a wonderful and at times overwhelming journey. I am indebted to many individuals without whom it would have been impossible to complete this unforgettable and challenging chapter.

I want to begin by acknowledging my supervisor, Dr. Vasilis Friderikos, for being a great role model. You have shown great flexibility and understanding whenever I needed support with issues both directly and indirectly relevant to my research. Your insight and the belief that a superior way to address specific areas of difficulties can always be found has guided and shaped my approach. It would not be an overstatement to say that my experience would have been very different indeed had I not had the fortune of having you as my guide.

I would like to express my heartfelt gratitude to Prof. Hamid Aghvami, Director of the Institute of Telecommunications for giving me the opportunity of pursuing a research degree. I have been very fortunate in having you as a mentor as I can state with some certainty that I would not have been in a position to present this work had it not been for your enabling encouragements and thoughtful guidance. Your ability to combine patience and critique whilst being encouraging in times of difficulties has supported me throughout the course of this study and for that I am truly grateful and forever in your debt.

I had the privilege of working with a friendly and brilliantly talented group of people in the Institute of Telecommunications. I would like to acknowledge all of my colleagues, in particular, Dr. Paul Pangelos for his moral support especially in times of crisis, thank you for being there. Special thanks goes to Dr. Toktam Mahmoodi for her technical feedbacks at different stages of my studies and for being a great friend, I admire your personality and highly value our friendship. I also wish to thank Dr. Andrej Mihailovic for all of the fruitful discussions we have had on technical subjects as well as political and philosophical ones.

I owe a debt of gratitude to Martin for his constant support and never ending faith in my abilities throughout the course of this work, especially providing invaluable tips at the final stages.

I would have never contemplated this journey if it weren't for my parents, who instilled in me the love of science and have been a beacon of hope throughout my life. You have shown unconditional and endless love, support and understanding at times when I really needed it. And last, but not least, to my brother Mani, whom his wise words gave me a lot of strength to march on despite the pitfalls, thank you.

Contents

Contents	6
Abbreviations	9
List of Figures	13
1 Introduction	15
1.1 Motivations	17
1.2 Contributions	18
1.3 Structure of the Thesis	19
2 Background Study	21
2.1 Introduction	21
2.2 Call Admission Control	22
2.2.1 CAC Classification	23
2.2.2 Objectives of Using CAC	25
2.3 Wireless Mesh Networks	27
2.3.1 WMNs Architecture	28
2.4 Mobility Support in All-IP Networks	32
2.4.1 Macro Mobility/Mobile IP	33
2.4.2 Micro Mobility	35

2.4.3	Proxy Mobile IPv6	35
2.5	CAC in Multi-hop Networks	39
2.5.1	Routing in WMNs	42
2.5.1.1	Routing Protocols for Mesh Networks	45
2.5.1.2	Routing Metrics	47
2.5.1.3	Dijkstra's algorithm	48
2.6	CAC in all-IP Networks with Mobility Support	50
3	Joint Revenue-based Call Admission Control and Routing	52
3.1	Introduction	52
3.2	Motivation and Background	55
3.2.1	Revenue-based CAC	55
3.2.2	Pricing in Telecommunications networks	58
3.3	Problem Description	59
3.4	System Model	61
3.4.1	Network Model	63
3.4.2	Routing Algorithm	65
3.4.3	Price and Revenue Function	67
3.4.4	CAC Algorithm	69
3.4.5	Complexity of the CAC Algorithm	70
3.4.6	Near-optimal Solution	71
3.5	Performance Evaluation	73
3.5.1	Simulation Scenario One: Evaluating the effect of pricing model	74
3.5.2	Simulation Scenario Two: Joint Revenue-based CAC and route assignment	77
3.6	Concluding Remarks	79

4 Admission Control Scheme for Proxy Mobile IPv6 Networks	80
4.1 Introduction	80
4.2 Background Study	84
4.2.1 A brief overview of QoS aware DRO and its Policy	84
4.3 Problem Description	88
4.4 System Model	90
4.4.1 Network Model	93
4.4.2 State Transition Diagram	95
4.5 Simulation Scenarios	98
4.5.1 First Scenario: All through the LMA	98
4.5.2 Second Scenario: Using RO schemes	101
4.6 Performance Evaluation	102
4.6.1 Blocking Probability of One Node	102
4.6.2 Blocking Probability Scenario I	104
4.6.3 Blocking Probability Scenario II	107
4.7 Concluding Remarks	111
5 Conclusions and Future Research	113
5.1 Concluding Remarks	113
5.2 Future Research	115
References	118
A Heuristic and Meta-Heuristic Algorithms	129

Abbreviations

3G	3rd Generation
AP	Access Point
AR	Access Router
BER	Bit Error Rate
BS	Base Station
BU	Binding Update
CAC	Call Admission Control
CDMA	Code Division Multiple Access
CN	Correspondent Node
CoA	Care of Address
DiffServ	Differentiated Service
DRO	Dynamic Route Optimisation
ETT	Expected Transmission Time
ETX	Expected Transmission Count

ABBREVIATIONS

FA	Foreign Agent
FMIPv6	Fast Mobile IPv6
GA	Greedy Algorithm
GCP	Greedy Call Packing
GSM	Global System for Mobile Communications
HA	Home Agent
HoA	Home Address
HMIPv6	Hierarchical Mobile IPv6
IETF	Internet Engineering Task Force
IntServ	Integrated Service
IP	Internet Protocol
LCoA	Local Care of Address
LMA	Local Mobility Anchor
LOS	Line Of Sight
MA	Mobility Agent
MAC	Medium Access Control
MAG	Mobility Access Gateway
MAP	Mesh Access Point
MANETs	Mobile Ad Hoc Networks

ABBREVIATIONS

MH	Mobile Host
MIPv6	Mobile IPv6
MKAR	Multiple Knapsack with Assignment Restrictions
MKP	Multiple Knapsack Problem
MN	Mobile Node
MT	Mobile Terminal
MSC	Mobile Switching Centre
NetLMM	Network-Based Localized Mobility Management protocol
NP	Nondeterministic Polynomial-time
P2P	Peer to Peer
PBA	Proxy Binding Acknowledgment
PBU	Proxy Binding Update
PDA	Personal Digital Assistant
PktPair	Per-hop Packet Pair Delay
P-LCoA	Proxy Local Care of Address
PMIPv6	Proxy Mobile IPv6
QoS	Quality of Service
QP	Queuing Priority
RA	Router Advertisement

ABBREVIATIONS

RO	Route Optimisation
RRM	Radio Resource Management
RSVP	Resource Reservation Protocol
RTT	Round Trip Time
SA	Simulated Annealing
SINR	Signal to Interference and Noise Ratio
SIR	Signal to Interference Ratio
SMDP	Semi-Markov Decision Process
TCP	Transmission Control Protocol
VoIP	Voice over IP
WCETT	Weighted Cumulative Expected Transmission Time
WiMAX	Worldwide inter-operability for Microwave Access
WIPPOA	Wireless IP Point Of Attachment
WLAN	Wireless Local Area Network
WMNs	Wireless Mesh Networks
WTP	Willingness to Pay

List of Figures

2.1	Infrastructure WMNs [1]	29
2.2	Client WMNs [1]	30
2.3	Hybrid WMNs [1]	31
2.4	Mobile IP Architecture [2]	33
2.5	The Architecture of PMIPv6 [2]	36
2.6	PMIPv6 Signalling [3][2]	38
2.7	CAC and routing in WMNs	43
3.1	Overview of joint Admission Control and Route Assignment Decision	63
3.2	Total Revenue gained vs Flows using the three random instances .	74
3.3	Blocking Probability vs Flows using the three random instances .	75
3.4	Topology of WMNs deployed in this work	76
3.5	Total Revenue gained vs Flows using the optimal pricing policy .	77
4.1	Direct and Triangular Routing in PMIPv6	86
4.2	Overview of the proposed resource management scheme for PMIPv6	92
4.3	Topology for Scenario I	94
4.4	Topology for Scenario II	95
4.5	Scenario I: z independent M/M/m/m queues	98
4.6	Class-based Admission Control Decision Making Process	100

LIST OF FIGURES

4.7	Blocking Probability vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points . . .	103
4.8	Total Blocking Probability vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points	105
4.9	Blocking Probability of non real-time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points	106
4.10	Blocking Probability of real-time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points	107
4.11	Total Blocking Probability vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points	108
4.12	Blocking Probability of non-real time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points	109
4.13	Blocking Probability of real-time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points	110

Chapter 1

Introduction

Evolution of wireless technologies over the past decades makes it almost impossible to imagine a world without mobile applications and services. The users are becoming more dependent on Internet and mobile services such as voice, text messages, Voice over IP (VoIP) and E-mail to get through their daily routines. With new devices and applications being developed, the number of users accessing multimedia rich services and the demand for ubiquitous “anytime and anywhere” communications is increasing. Therefore, evolution of wireless and device technologies proliferate wireless resource requirements. Multimedia applications require strict Quality of Service (QoS) support. However, QoS provisioning for the wide array of existing and new services is affected by the scarcity of wireless resources. Efficient design schemes are required to provide appropriate Radio Resource Management (RRM) policies.

The promise of advanced multimedia services by the network providers to the users has become closer to reality with the advancement of wireless devices such as smart phones. This introduced a new avenue for revenue generation and new

challenges for the network providers. Given the huge load of multimedia traffic on the network, providing QoS whilst simultaneously increasing the revenue is an extremely difficult job.

There has been a vast amount of interest in extending the coverage area of IP networks through multi-hop communication to areas where providing IP connectivity is almost impossible such as rural areas. Wireless Mesh Networks (WMNs) are a promising technology due to certain capabilities such as extending the coverage area of IP networks and low cost of deployment and maintenance. However, many challenges exist in designing an efficient QoS mechanism for such multi-hop based networks, which are thoroughly discussed in this thesis. Moreover, the Internet is designed for static networks and IP does not natively support mobility. The Internet Engineering Task Force (IETF) has been working on different protocols in order to provide mobility support for the Internet. Its work on the subject has led to the birth of many protocols such as macro and micro mobility solutions that are further explained in the following chapter.

Leveraging limited wireless resources to prevent congestion and support QoS requirements by the users' needs Call Admission Control (CAC), which is a part of RRM scheme. CAC is the decision process of accepting or rejecting a user's access to the network [4] based on the available network resources. The objectives and design criteria of CAC mechanisms have evolved alongside the wireless networks. Various generations of wireless systems required alternative QoS support leading to addition of new dimensions, hence increasing the complexity in CAC design schemes. This has been comprehensively outlined in Chapter 2.

1.1 Motivations

An efficient CAC mechanism bases its decision with respect to the available bandwidth and congestion level in the network. This depends on network architecture as well as design objectives of the CAC mechanism. Future IP based networks have a multi-hop wireless infrastructure similar to WMNs and are expected to provide mobility support. This thesis revolves around designing CAC mechanisms based on the challenges that salient features in future multi-hop based networks such as WMNs and micro mobility solutions such as Proxy Mobile IPv6 (PMIPv6) impose.

In WMNs common nodes or links between the backhaul paths limit the amount of traffic that can be admitted into the network. Calculating available bandwidth requires an accurate picture of the network topology, therefore CAC decision in such networks is closely coupled with route selection. The access link is another area with high probability of becoming congested, as it connects the mesh clients to the Access Points (AP). The CAC design complexity lies in assigning mesh clients to APs that have enough resources in both the access and backhaul links.

PMIPv6 belongs to a family of micro mobility solutions proposed by IETF. In spite of eradicating many of the issues that were associated with other macro and micro mobility solutions such as handover delays and excessive signalling over the wireless link, QoS provisioning in PMIPv6 still remains a challenge. The presence of Mobility Agents (MA) overutilises certain paths around the MA and create a bottleneck effect. Lack of appropriate CAC schemes at the MAs and underutilising the path that do not contain MAs reduces the overall capacity of the network. CAC mechanism must address the issue of resource management in the presence of MAs with the aim of increasing the throughput of the network.

1.2 Contributions

This thesis has made the following contributions:

- To address the issue of QoS provisioning in WMNs through admission control, a joint CAC and route selection algorithm is formulated as an optimisation problem that aims at maximising the revenue of the network operator. Effects of applying different pricing strategies in the network are explored; an optimal pricing policy that considers the complexity of the problem and the connection dropping probability is derived. Two sub-optimal solutions for the joint CAC and route selection algorithm are found through extensive simulation studies in order to provide a solution for the proposed optimisation problem in real time.
- To counter the problem of bottleneck effect in presence of MA/s in PMIPv6, a novel class-based CAC scheme is proposed. PMIPv6 is modelled as a tandem queuing network with two types of arrival process, and the performance of each node within the network was analysed by a 2-dimensional Markov chain. The main novelty of this approach is division of the available resources to multiple resource units and design of the admission policies that gives more priority to real-time traffic. It has been shown that in addition to significantly reducing the total blocking probability, per class blocking probability was also reduced in comparison with the case in which the proposed CAC scheme was not used.
- Finally, to eradicate the bottleneck effect caused by triangular routing, a modified version of a previously proposed RO scheme is incorporated that

operates alongside the proposed admission control scheme. The network performance was improved through applying the proposed CAC scheme and routing non real-time away from the MA. The joint approach provided benefits over the scenario where no RO was considered and only the proposed CAC scheme was implemented.

The publications that are related to the contributions made in this thesis are:

1. Nika Naghavi, Vasilis Friderikos, Toktam Mahmoodi, Hamid Aghvami, “Admission Control Scheme for Proxy Mobile IPv6 Networks,” IEEE International Conference on Communications, ICC’13, Budapest, Hungary, 9-13 June 2013.
2. Nika Naghavi, Vasilis Friderikos, Hamid Aghvami, “Joint Revenue-based Call Admission Control and Routing in Wireless Mesh Networks,” IEEE International Conference on Communications, ICC’11, Kyoto, Japan, 5-9 June 2011.
3. Nika Naghavi, Vasilis Friderikos, “A Batch Based Call Admission Control Algorithm for Wireless Mesh Networks,” London Communications Symposium, LCS2009, London, UK.

1.3 Structure of the Thesis

The remainder of this thesis is structured as follows:

The following chapter sets out a summary of the related literature with the aim of building an overview of the relevant work in CAC mechanisms for WMNs and PMIPv6. The main contributions of the thesis are embodied in Chapter 3 and 4.

In Chapter 3, joint association of mesh clients with APs alongside route selection from APs to the gateway is modelled as an optimisation problem of maximising revenue of the network provider. Chapter 4 extends the issue of admission control design for future networks by considering the effect that presence of MAs have on available resources. A class-based CAC scheme which assigns more resource units to real-time traffic is proposed. An RO scheme on top of the proposed CAC scheme is implemented which aims at increasing available resources and enhancing the overall throughput of the network. Finally, in Chapter 5 concluding remarks and potential avenues for future research are discussed.

Chapter 2

Background Study

2.1 Introduction

In this chapter, the background and the state of the art research with respect to the work presented in the remainder of this thesis is presented. The first part of this background study surveys a full overview of Call Admission Control (CAC) and its evolution over the three generations of wireless systems and beyond. With the advances of wireless technology and the growing interest in data and multimedia services, new challenges are imposed in the design of the CAC schemes. The central theme of this thesis is based on designing an optimal CAC scheme in multi-hop based networks such as Wireless Mesh Networks (WMNs) and in networks with mobility support like Proxy Mobile IPv6 (PMIPv6). To this end, section 2.3 outlines a detailed description of WMNs. Section 2.4 gives an overview of mobility support mechanisms in all-IP networks and provides a thorough explanation on how PMIPv6 networks operate. Finally, after separately covering the background on CAC, WMNs and PMIPv6 networks, the issues and

challenges involved in designing a CAC scheme for multi-hop networks and all-IP networks with mobility support are discussed and the related literature is stated.

2.2 Call Admission Control

Radio Resource Management (RRM) is one of the key roles of QoS provisioning mechanisms as it can have an impact on the user or the network performance. Radio resources can be managed using an array of schemes but in general they can be classified in three sets [5]. Frequency/time resource allocation schemes are in the first set which includes channel allocation, scheduling, transmission rate control, and bandwidth reservation schemes. Schemes in the second set are based on controlling the transmitter power of the terminals and the Base Stations (BS), power allocation and control schemes are part of this set. Schemes focused on controlling the access port connection such as CAC, BS assignment, and handover algorithms belong to the third set. CAC is the process of controlling which user might or might not access the network based on a predefined criteria and can be defined as a strategy for accepting, rejecting and negotiating users' call requests [4][6]. An admission controller accepts a new call request if enough resources are available to meet the QoS requirements of the new call without violating the QoS of the already admitted calls [7].

CAC has been extensively studied in wireline networks and has been a prominent field of research in wireless networks. In the first two generations of wireless systems, a great deal of effort was put into designing an efficient CAC for voice services. However, majority of the proposed CAC algorithms for the first and second generations were not suitable for the third generation (3G) and beyond networks as they were only designed for a single service environment. Sophisti-

cated design schemes are required to cope with the QoS requirements of data and multimedia services supported by 3G and beyond networks.

2.2.1 CAC Classification

Classifying CAC schemes is a complex and subjective task. Different variables are involved in CAC schemes but based on the design option used, schemes can be divided into different classes [5]. According to [8] CAC can be classified into the following categories:

- Centralised and Distributed Approaches:

The CAC algorithm can either be categorised as centralised or distributed depending on where the decision of CAC algorithm is executed. In a centralized fashion, the CAC algorithm is performed in a central site (i.e. the Mobile Switching Centre (MSC)) rather than the local cell. This dictates the transformation of the information from each cell to the central site, where the CAC decision is made remotely. In the latter approach, as the name suggest, the decision is made locally at the base station of each cell. A distributed CAC can further be partitioned into two different approaches: a collaborative and a non-collaborative approach. In the former case the information regarding the resource reservation and admission control is shared and exchanged between the neighbouring cells, whereas in the non-collaborative approach the CAC decision is made based on the local information in the cell. A more precise CAC decision can be made using the collaborative approach but it requires more signalling which leads into communication overhead.

- Traffic-Descriptor-Based and Measurement-Based Approaches:

In the traffic-descriptor approach, the knowledge about the incoming calls' traffic pattern is assumed to be known. In order to make a decision, the CAC algorithm establishes the amount of resource usage by summing up all the resources used by the ongoing calls and the resources to be used by incoming calls. An incoming call is accepted if the sum is less than a pre-defined threshold and is rejected otherwise. This approach is unadventurous in the sense that it may result in rejection of ongoing calls that might not use the maximum amount of resources as predicted in the descriptor. This approach is widely deployed in the hard-capacity GSM networks. In the measurement-based approach the traffic pattern information is achieved through measuring the call characteristics. The CAC decision is then made based on the state of the network and has the advantage of being made dynamically. Most CAC algorithms deployed in Code Division Multiple Access (CDMA) networks are measurement based where QoS criterion of the ongoing call is satisfied by measuring the Signal to Interference Ratio (SIR).

- Based on the Granularity of Resource Control:

Different CAC algorithms can be differentiated based on the granularity of the resource control. Three different criteria in categorising the CAC are: the type of information, the spatial distribution and the way the information is organised and manipulated. The first criterion is normally considered as the resource usage of mobiles. The second criterion or the position and movement of the users can be either uniform or non-uniform. The information can be on aggregate of flows (group of mobiles) or per flow (individual mobile) and the decision based on this information can be performed on

either.

However, more classes based on the design option can be mentioned. For instance CAC can be classified based on number of services considered. As mentioned earlier single-class (i.e. voice) CAC schemes received a great deal of attention in the first two generations of wireless systems [5]. Multiple-class CAC schemes are more relevant for 3G and beyond networks in order to provide QoS support for voice, data and multimedia services. Another classification is based on whether the CAC was designed for the uplink [9], the downlink [10] or for both links jointly [11]. Some services such as voice connections require symmetric bandwidth in both uplink and downlink; therefore in most cases a disjoint CAC scheme based on either links suffices. However, services like streaming might need greater bandwidth in downlink than uplink and a joint approach is more suitable. Finally, optimisation can be another design option for CAC schemes. Optimal CAC schemes are more desirable but they are not achievable in most realistic scenarios especially where the problem is complex due to its size. In such scenarios heuristic and meta-heuristic techniques are incorporated to achieve suboptimal CAC schemes.

2.2.2 Objectives of Using CAC

The main reason for using CAC schemes is to guarantee QoS parameters. Ahmed at [5] has summarised main objectives of using CAC, some of which are:

- Controlling the signal quality:

Providing a minimum signal quality to the new and admitted users is the main aim of some CAC schemes. An appropriate metric for measuring the signal quality is the Signal to Interference and Noise Ratio (SINR) as there is

a direct relationship between the SINR and the Bit Error Rate (BER). The SIR is usually used instead of SINR in interference-limited networks. The aim of majority of CAC schemes that control the signal quality is to achieve either a minimum SIR or SINR. Different approaches can be incorporated to achieve a minimum signal quality. The SIR or SINR based CAC schemes as proposed in [9] and [12], are one of these approaches where incoming calls are admitted as long as the measured SIR or SINR by admitting this new call is higher than a (SIR or SINR) threshold. Controlling the number of users in a cell or a network can be another approach in maintaining the signal quality. Finally, the transmitted/received power is a good measure of signal quality offered to the ongoing calls as well as the new calls. Transmitted/received power was widely used as admission criterion in CDMA networks.

- Controlling handover failure probability:

Dropping an admitted call is more annoying than blocking a new call from the user's perspective, therefore controlling handover failure probability have been the objective of some CAC schemes. Controlling the handover probability generally is anonymous with prioritising handover calls over new calls and consequently there is a trade-off between reducing dropping probability and new call blocking probability. Different approaches have been suggested in literature to control the handover probability, which can be broadly differentiated as follows [13]: I) Guard Band Schemes, where certain amount of channels are reserved or kept for handover calls II) Queuing Priority (QP) Schemes where different priorities are assigned to either handover or new calls.

- Controlling packet-level QoS parameters:

Packet-switched networks introduced new parameters that must be taken into account as an admission criterion, for instance such networks can experience long delay jitter or packet delay when the network is overloaded. Also, supporting multimedia traffic over packet-switched networks requires stringent packet-level QoS parameters such as packet delay, jitter, and loss [14]. Therefore, controlling packet-level QoS parameters have been the main objective of many CAC schemes.

- Service/class prioritisation:

Assigning higher priority to a certain class of traffic has been the objective of many CAC schemes. This approach is adapted to differentiate between the classes of traffic in the network. The differentiation between classes of traffic can be made based on different underlying objective such as potential reward level to the network operator, sensitivity with regards to delay and others. Varying approaches may be used to provide prioritised CAC schemes such as using different threshold levels (number of users, power level, resource units and etc) for each class of traffic.

2.3 Wireless Mesh Networks

WMNs will play a significantly important role in the next generation Internet by extending IP connectivity to regions which are unreachable by any single access technology [15] (Chapter 1), [16][1] (Chapter 1). They can provide long coverage range without affecting the channel capacity through mesh-style multi-hop transmission over short distances which facilitates less interference between the nodes

and more efficient frequency re-use. Also, they can provide non-Line Of Sight (LOS) connectivity for the users without direct LOS through multi-hop transmission. Deploying cable in some areas such as remote valleys and rural areas can either be really costly or impossibly hard and WMNs can be easily deployed in such areas. Self-forming, self-healing and self-organising nature of WMNs have many associated benefits such as low upfront investment, easy network deployment and maintenance, and robustness. WMNs have the advantage of being readily implemented on existing technologies and in general are considered to be a promising technology for the next generation of wireless networks.

WMNs are multi-hop wireless networks consisting of mesh routers and clients. Each node can operate as both a host and a router; forward packets to the destination nodes, which are not within the transmission range of the sender nodes. Mesh routers are wirelessly inter-connected in order to build the backhaul structure; they have minimal mobility and do not have many restrictions on the power consumption. The following formal definition is presented in [17]: “A wireless mesh network is a packet-switched network with a static wireless backbone”, from which it can be concluded that the wireless backbone topology is fixed. Mesh clients can be from a wide range of devices such as desktops, laptops and phones, and they can be mobile or stationary.

2.3.1 WMNs Architecture

Wireless mesh routers can mainly provide two sets of functionalities in a network. They can extend the network coverage area and act as repeaters to form the mesh backbone. Gateway functionalities is the second set and wireless mesh routers that are connected to the Internet are referred to as Gateways. Additional

routing functionalities that facilitate mesh networking [16] [1] (Chapter 1) are also available on the mesh routers. Mesh clients can communicate with other clients across the Internet via the gateway. They also benefit from mesh networking functionalities and as a result can operate as a router. Mesh clients do not have the bridge or gateway functionalities.

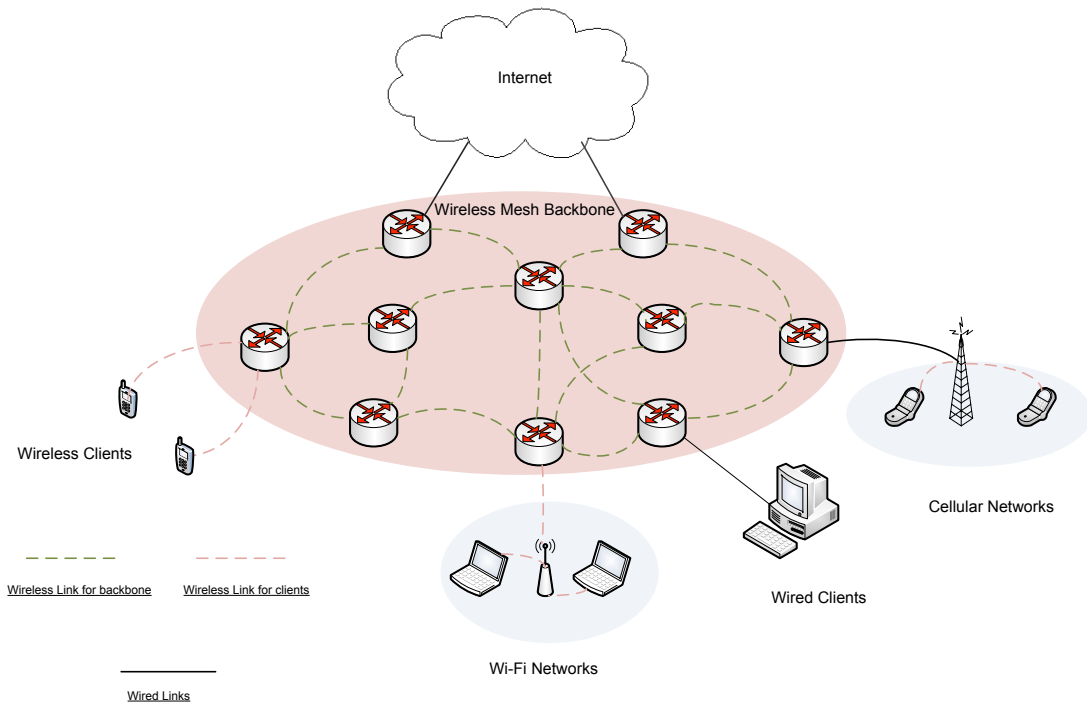


Figure 2.1: Infrastructure WMNs [1]

The basic architecture of the WMNs can be classified into three groups based on the network topology [16] [1] (Chapter 1):

1. Infrastructure / backbone WMNs:

In infrastructure WMNs, the network has multiple tiers in which WMNs clients form the lowest in the hierarchy. Mesh routers form an infrastructure (or WMNs backbone) for mesh clients to connect to them. This backbone can be built using

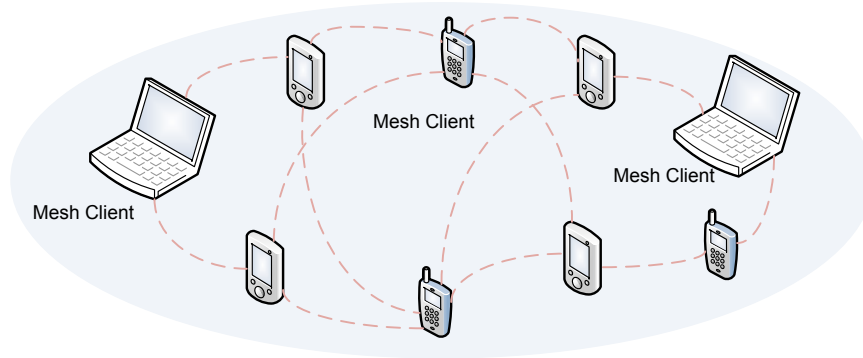


Figure 2.2: Client WMNs [1]

different radio technologies but the most commonly used one is IEEE 802.11. The mesh routers are normally equipped with multiple interfaces to connect the other networks such as Internet, Cellular, Wi-Fi and WiMAX. They are responsible for self-organising and maintaining the backbone. Each mesh router use two types of radio; one for backbone communication and one for communication with the mesh clients. As it is shown in the Figure 2.1, mesh routers using their gateway functionality can be connected to the Internet and form the gateways. The mesh clients with Ethernet interface can connect to the mesh routers via Ethernet links, clients with a different radio technology can be connected to the mesh routers via their BS which has an Ethernet interface, as shown in Figure 2.1. Overall this approach facilitates a backbone for conventional clients and provides the integration of WMNs networks with existing wireless networks.

2. Client WMNs:

Consists of only mesh clients and no mesh routers. The client WMNs are very similar to the ad-hoc networks [15], where a small group of users can communicate with each other through multi-hop techniques. A packet destined to a node hops through multiple nodes to reach the destination. The basic architecture is shown in Figure 2.2. Client devices act as both routers and hosts and each node is at

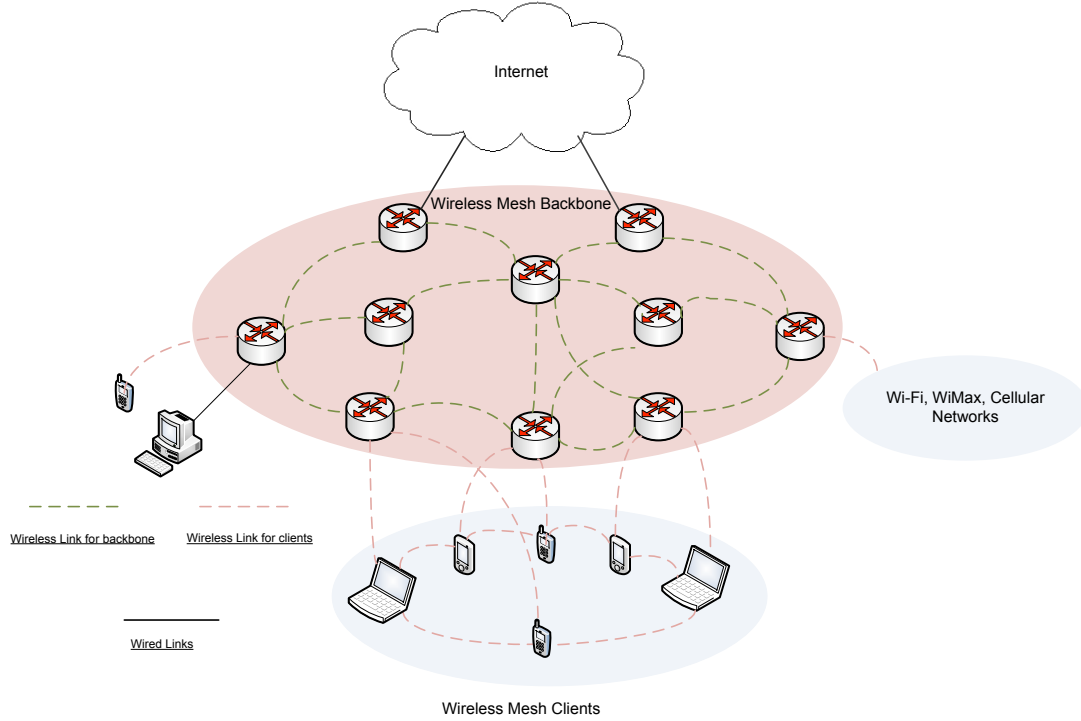


Figure 2.3: Hybrid WMNs [1]

the same level of its peers. Usually one type of radio is used in order to form client WMNs.

3. Hybrid WMNs:

This architecture combines the infrastructure/backbone WMNs with client WMNs; therefore mesh clients can access the network through either meshing with each other or through the mesh routers. This has the benefit of both connectivity to other networks such as Internet, Wi-Fi alongside better connectivity and coverage inside the WMNs through routing capabilities of the clients. The basic architecture is shown in Figure 2.3.

2.4 Mobility Support in All-IP Networks

Next generation of mobile networks are expected to support a variety of mobile devices such as smart phones, Personal Digital Assistants (PDAs), and laptop computers over IP based mobile networks. All-IP networks that are a combination of Internet and telecommunications network, rely entirely on IP, from the mobile station to the gateway towards the Internet and in both wired and wireless parts of the network. Such a network can be described as a set of Wireless IP Point of Attachments (WIPPOA) that is connected to an IP backbone, with a gateway that is connected to the Internet. One of the most desirable features of cellular networks from the user's point of view was or is the ability to move and be mobile whilst being connected to a communication network. Unfortunately, IP does not natively support mobility and IP addresses (IPv4 and IPv6) were mainly designed to be used as locators and identifiers at the same time [2]. IP addresses are locators which specify how to reach a certain node (i.e. the network interface) that is using a specific destination address. IP addresses are also used as identifiers by upper-layer protocols such as the Transmission Control Protocol (TCP) to distinguish the endpoints of a communication channel. Supporting mobility requires separating the identifier role from the locator role. From the identification perspective the address of a node should never change but from the location point of view when a movement happens the IP address must change in order to show the current location of a node and the subnet to which it is attached to. This problem was addressed by Mobile IP [18][19], which is detailed in the next subsection.

Mobility management in the context of IP is managing the change of attachment of the Mobile Node (MN) and is further divided into Micro/Macro-mobility. A domain is a large wireless access network under a single authority, which may

include varied access networks. The movement of MNs from one domain to another domain (global) and within one domain (local) are referred to as Macro and Micro mobility respectively, this is illustrated in Figure 2.4.

2.4.1 Macro Mobility/Mobile IP

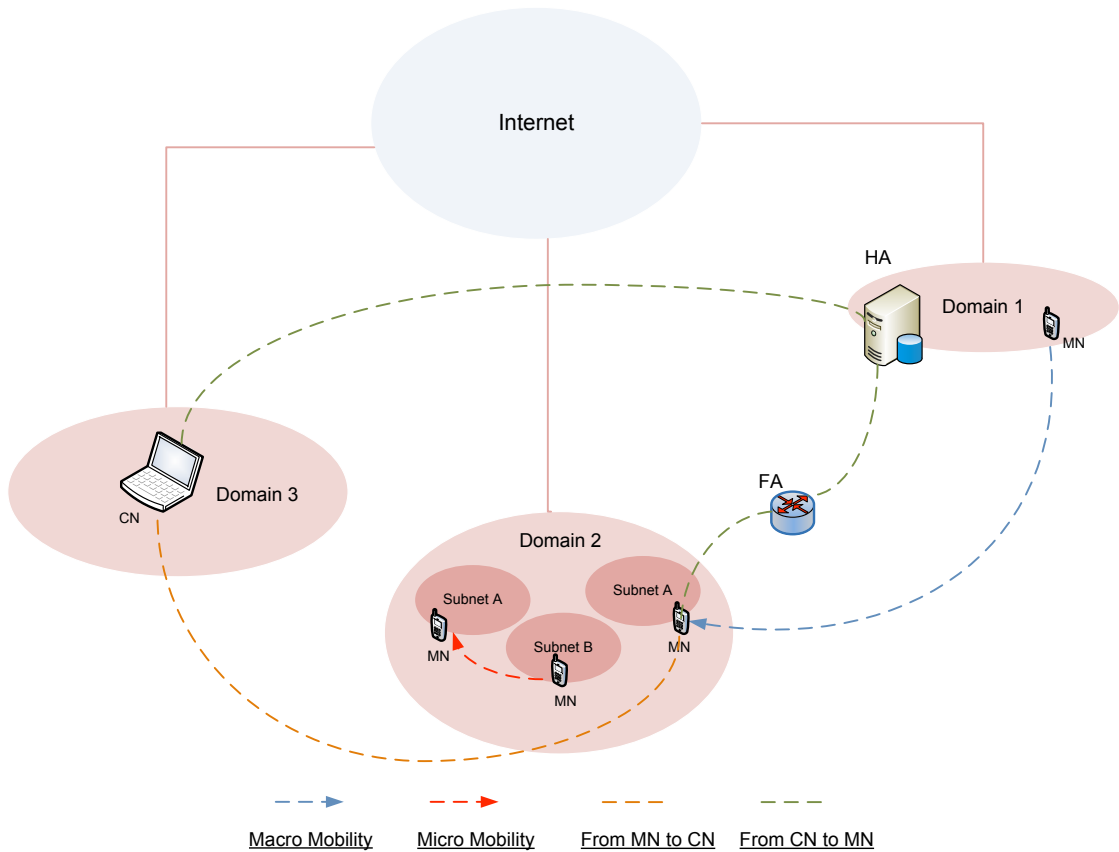


Figure 2.4: Mobile IP Architecture [2]

Mobile IP [18][19] was proposed by IETF to provide IP connectivity to the MN and it introduced three new entities: MN which is the same as Mobile Terminal (MT) or Mobile Host (MH)), Home Agent (HA) and Foreign Agent (FA) (only used in Mobile IPv4)[20]. The HA serving an MN must be placed at the home

2.4. Mobility Support in All-IP Networks

network, an FA may be present in the network that the MN is visiting. The IETF also developed IP-layer solutions for the problem of MN mobility in Mobile IP. The solutions were based on providing the MN with two sets of addresses: 1. A permanent address, called the Home Address (HoA) which is to be used as an identifier and 2. A temporal address, referred to as the Care of Address (CoA) that is used as a locator [2]. The location management of the MN as it moves across the Internet is performed by the HA and each MN has a home address which the HA recognises. When an MN moves into a new IP domain, it obtains a Local Care of Address (LCoA). Then it sends a Binding Update (BU) to the HA to bind the new CoA at the present network with its home IP Address. Any packets sent by a Correspondent Node (CN) to an MN are intercepted by the HA. The HA then tunnels the traffic to the MN's CoA. If there is an FA on the visiting network, it receives the tunnelled packets and forwards them to the MN. The same does not hold for the reverse direction, as the MN directly sends packets to the CN. This process is shown in Figure 2.4. The movement of MNs from one domain to another domain and within one domain are referred to as Macro and Micro mobility respectively, this is also illustrated in Figure 2.4.

A large handover delay is associated with Mobile IP, this is due to signalling exchanges between the MN and the HA. In order to tackle the large handover latencies in Mobile IP a new family of solutions called micro mobility solution were proposed to provide localised mobility support. This is explained in the next subsection.

2.4.2 Micro Mobility

Micro mobility solutions aim at reducing the handover detection and performance latencies inside a domain. In Mobile IP every time an MN changes its WIPPOA, signalling exchanges are needed between the MN and HA. This creates an overload of signalling and delays to the home network as an MN makes frequent moves locally or within subnets of one domain. Micro mobility solutions were proposed to reduce signalling overload and lower handover latencies during movements within one domain. A local HA or an HA closer to the MN is used during mobility in the local domain. The MN's movement within one domain is transparent to the HA and the rest of the network and is managed by the local HA. This results in faster signalling exchanges and reduced signalling overload to the home network [20].

Various micro mobility protocols such as Cellular IP [21], Fast Mobile IPv6 (FMIPv6) [22], and Hierarchical Mobile IPv6 (HMIPv6) [23] and others [24] were proposed. All of these protocols as well as Mobile IP are MN based and require an upgrade in the protocol stack of MNs which leads to more complexity. Moreover, the software stack change in MNs may not be compatible with all macro mobility protocols. A thorough description of Proxy Mobile IPv6 (PMIPv6) operation, which is a network based mobility management solution is provided in the following subsection.

2.4.3 Proxy Mobile IPv6

There has been a great interest in micro mobility solutions that relocate mobility procedures from the MN to the network, which has led to the birth of

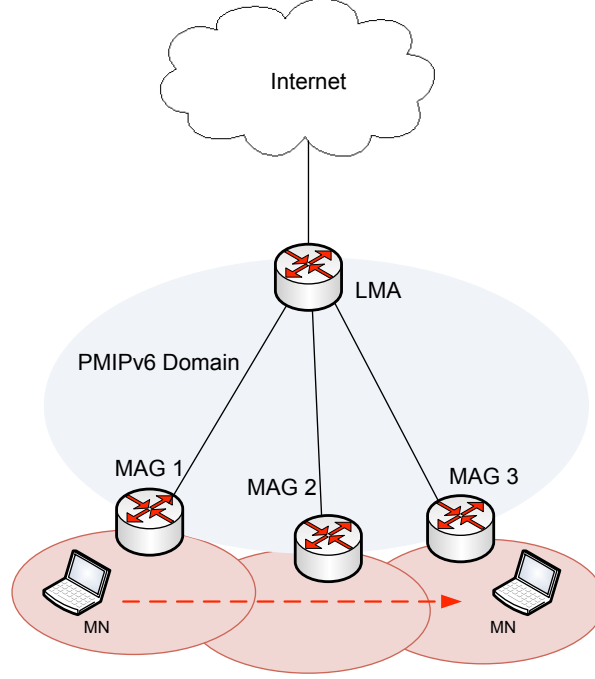


Figure 2.5: The Architecture of PMIPv6 [2]

Network-Based Localized Mobility Management protocol (NetLMM) [25]. The main advantage of NetLMM is that network operators are able to provide local mobility support without the need of additional software changes by the MN. PMIPv6 is a network based mobility management protocol that enables IP mobility without involving the MN [3] and was standardised by the IETF in order to provide NetLMM in IP networks. Its advantages over the existing protocols range from minimising the handover latencies, to reducing the overheads such as signalling over the wireless link and non-complex deployment. In PMIPv6, the MN does not participate in any of the mobility-related signalling and the serving network is in charge of the MN's mobility management. The two main logical entities that provide a network-based mobility management support in PMIPv6 are: 1. Local Mobility Anchor (LMA) and 2. Mobility Access Gateway (MAG). The LMA is similar to the HA in (Mobile IPv6) MIPv6, in the sense that it has a

binding cache memory for all the registered MNs and it's the entity that manages the MNs BUs. Multiple LMAs may exist in a PMIPv6 domain, where each of them is responsible of serving a different group of MNs. The MAG is usually run on the Access Router (AR) of the MN. The MAG's main duty is to perform mobility-related signalling as well as detecting and performing handovers on behalf of all the MNs that are attached to its access link [3][26]. The functional entities of PMIPv6 are shown in Figure 2.5.

When an MN enters the PMIPv6 domain, it attaches itself to an access link of a MAG. The MAG then has to authorise whether the MN is allowed to use network based mobility management service. Once the MAG has authenticated the MN's access, it will send a Proxy Binding Update (PBU) to the LMA. If the sender is established to be a trusted MAG, the LMA accepts the PBU and in return sends a Proxy Binding Acknowledgment (PBA) to the MAG that includes the home network prefix. The LMA also creates a binding cache entry and sets up a bidirectional tunnel to the MAG. This tunnel is different from the tunnel established in MIPv6, it is only established between the LMA and the MAG [26]. A Router Advertisement (RA) message is sent by the MAG to the MN and the MN configures its own home address. All traffic originated or sent from the MN is routed to the LMA through the bidirectional tunnel and are then forwarded to the destination CN by the LMA. Also any traffic destined to the MN is received by the LMA and is forwarded to the serving MAG through the tunnel. The traffic is then sent by the MAG to the MN through the access link. If the MN changes its point of attachment, the MAG detects this detachment or movement by the MN and sends a signal to the LMA. Upon receipt of this signal, the LMA goes to a wait mode and allows the MAG on the new access link to send a BU. The new MAG sends a PBU to the LMA and updates the location of the MN to the LMA. The same prefix through the RA message is sent by the new MAG to the

2.4.Mobility Support in All-IP Networks

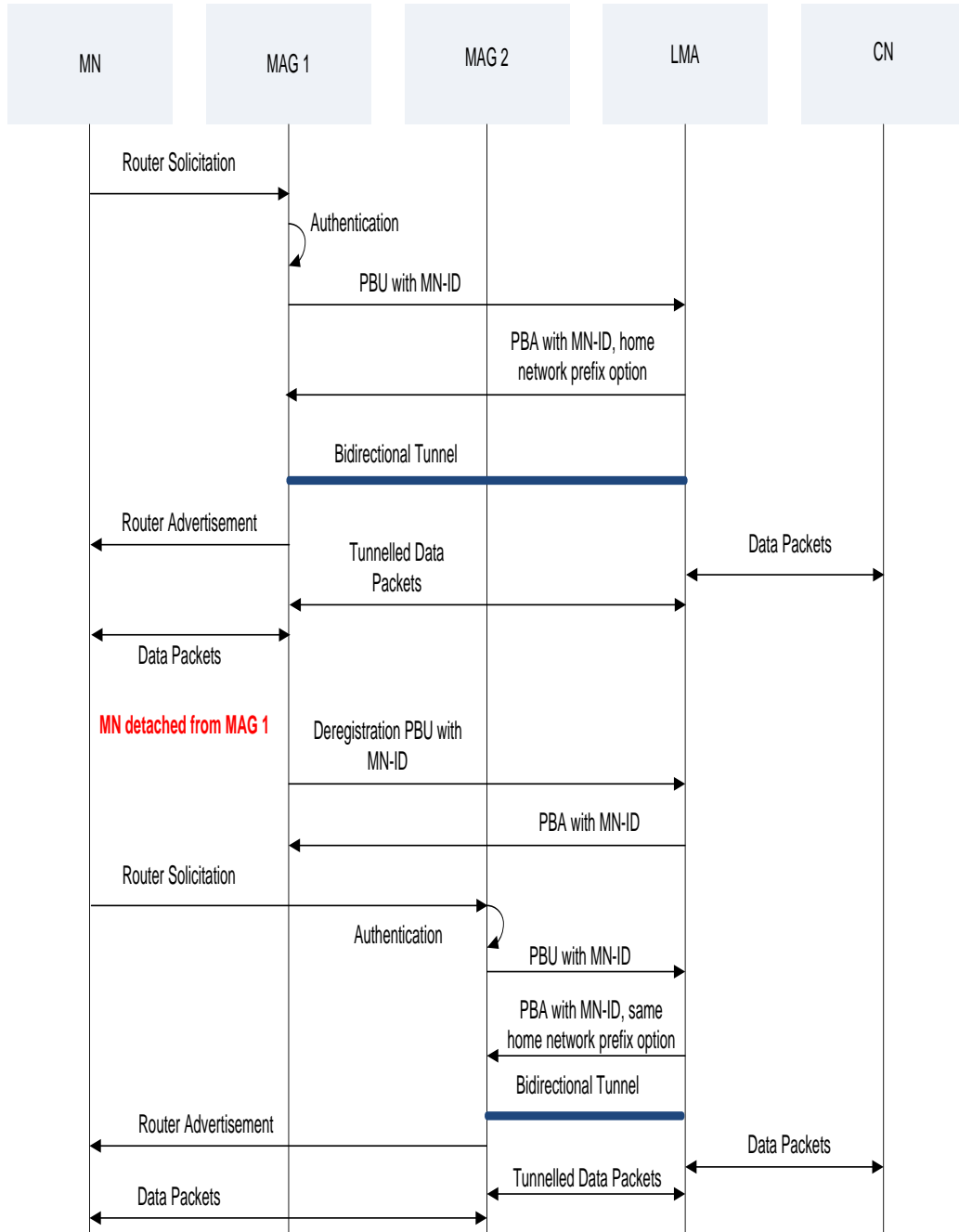


Figure 2.6: PMIPv6 Signalling [3][2]

MN. As a result the MN will not change the first address configured when it entered the PMIPv6 domain and the MN is kept free of any mobility signalling [3] [27]. Figure 2.6 shows the signalling when an MN enters the PMIPv6 domain and when it changes its point of attachment.

2.5 CAC in Multi-hop Networks

A large body of work has been published on admission control schemes for wireless cellular networks [5], and the majority of suggested mechanisms were focused on either providing a minimum SIR or reducing the probability of handover failure. However, admission control schemes proposed for cellular networks can not be extended to multi-hop networks. MNs can directly communicate with the BS in cellular networks or in all the single-hop networks such as WLANs and the admission control is based on the availability of resources at one single node. In multi-hop networks many new factors are involved in designing a CAC scheme which makes it a challenging task. Some of these challenges are due to the decentralised nature of multi-hop networks, power or energy constraints of the nodes, interference from the neighbouring and hidden nodes, contention for accessing the wireless channel, routing issues and lack of infrastructure [5] in such networks. Therefore, an optimal CAC scheme in multi-hop networks should jointly consider some of the above mentioned factors.

Designing a CAC mechanism for Mobile Ad Hoc Networks (MANETs) is a complex task due to many challenges posed by such networks [28]. For instance in MANETs the wireless channel is unreliable due to interference from other transmissions, shadowing and multi-path fading effects. Also there is a lack of centralised control as the network topology dynamically changes. Different CAC

schemes were proposed for multi-hop MANETs, in [29] a bandwidth reservation-based CAC is proposed for multi-hop MANETs which works in conjunction with an on-demand QoS routing protocol. The available bandwidth is found through calculating saturation throughput of a node. A contention-aware CAC is introduced in [30] to support QoS in multi-hop MANETs. In their work, each node transmits only at times when the medium is idle, the admission decision is made based on the bandwidth of the flow and the amount of available bandwidth both at the node and the neighbouring nodes. A comprehensive survey of the admission control mechanisms designed for 802.11-based multi-hop MANETs is given in [28].

It can be stated that WMNs are considered as a type of ad-hoc networks based on their architecture as depicted in section 2.3, but they have additional capabilities and different characteristics compared to the ad-hoc networks. WMNs intend to diversify the ad-hoc networks capabilities and therefore in some literature ad-hoc networks are categorised as a subset of WMNs [16]. One of the main differences is that WMNs benefit from a fairly stable infrastructure and topology, while in ad-hoc networks connectivity and topology depends on the mobility rate of the users. Routing in ad-hoc networks can be done through the end users, whereas in WMNs routing is performed by the mesh routers [16]. Therefore, in WMNs the load on the end-user devices is much less than ad-hoc networks and lower energy consumption in such devices is achievable making them an attractive choice for mobile and energy constrained end-users. Moreover, employing WMNs can improve the overall capacity of the network; as stated earlier mesh routers use two different radios for the access and routing functionalities, whereas in ad-hoc networks the two functionalities are performed on the same channel. For the reasons stated above CAC algorithms developed in ad-hoc networks might not be an appropriate choice in the WMNs.

The available bandwidth in WMNs depends on the underlying network technology with data rates as high as 54 Mbps possible [17]. Some of the challenges involved in designing an optimal CAC scheme in WMNs are similar to the MANETs. The absence of a central coordination authority in WMNs creates many complexities, for instance multiple flows which are sharing the same node or link, may simultaneously send probe messages in order to check the availability of the resources [31]. Also, the contention range of a node maybe beyond its communication range which can have an impact on the transmission of all neighbouring nodes. Therefore, it can be stated that estimating the amount of available bandwidth at mesh clients or routers is affected by the amount of traffic in both the transmission and sensing range of a node, node mobility and the state of the wireless links [31]. Hence, making the task of designing an optimal CAC scheme a complex one.

Vast amount of research has been carried out on the subject of CAC mechanisms in WMNs. A joint optimisation of association, backhaul routing and bandwidth allocation in WMNs is defined in [32]. The objective of this work is to maximise the network throughput as well as assurance of a fair bandwidth allocation amongst all the mesh clients, considering the access and backhaul links capacity and taking the wireless interference into account. The author first formulates a fractional association and multipath routing, upon addition of single-path routing and integral association constraints the problem becomes a mixed integer non-linear programming optimisation problem. To tackle the hardness of the problem, the logical topology construction and bandwidth allocation are considered separately. The wireless interference is modelled using a clique approximation, in order to improve the complexity of constructing the wireless interference constraints. The performance loss due to clique approximation is then recovered, using a scheduling algorithm that employs time slot reuse. A contention-aware admission control for multi-hop WLANs is suggested in [33], in which a con-

tention graph is utilised to model the wireless interference in the network. In the proposed scheme, a maximal clique is defined as a sub-graph that does not belong to any other sub-graph or cliques. A new flow can only be admitted if the total traffic on each maximal clique is not more than the available capacity in that clique. The capacity of a maximal clique is estimated as the saturation throughput. A revenue based connection admission control and routing, which is modeled as a Semi-Markov Decision Process (SMDP) is addressed in [34]. It assumes k pre-computed paths for each origin-destination. A resource reservation scheme for classes or services possessing higher priority is embedded. However the admitting probability is used as the average reward criteria and the work does not delve into the effect of pricing model on the performance of the system such as total revenue gain or connection blocking probability. An admission control algorithm based on the connections' rate and delay is proposed in [35]. In the proposed CAC scheme, first a tree-based topology connecting wireless backhaul nodes to the wired gateway is constructed, the admission decision is then carried out with the objective of revenue maximising whilst taking into account the rate and delay of the connections.

2.5.1 Routing in WMNs

Routing in multi-hop networks can be used as a tool to control congestion and improve the overall capacity of the network as such they can have an impact on the CAC design scheme and decision. As opposed to the single hop networks, in WMNs the admission of the traffic depends on the amount of available capacity in both the access links and the backhaul links. Figure 2.7 illustrates how routing and CAC are interrelated in WMNs networks, where routes were set up previously. Upon arrival of a new traffic, the admission decision based on the residual capacity

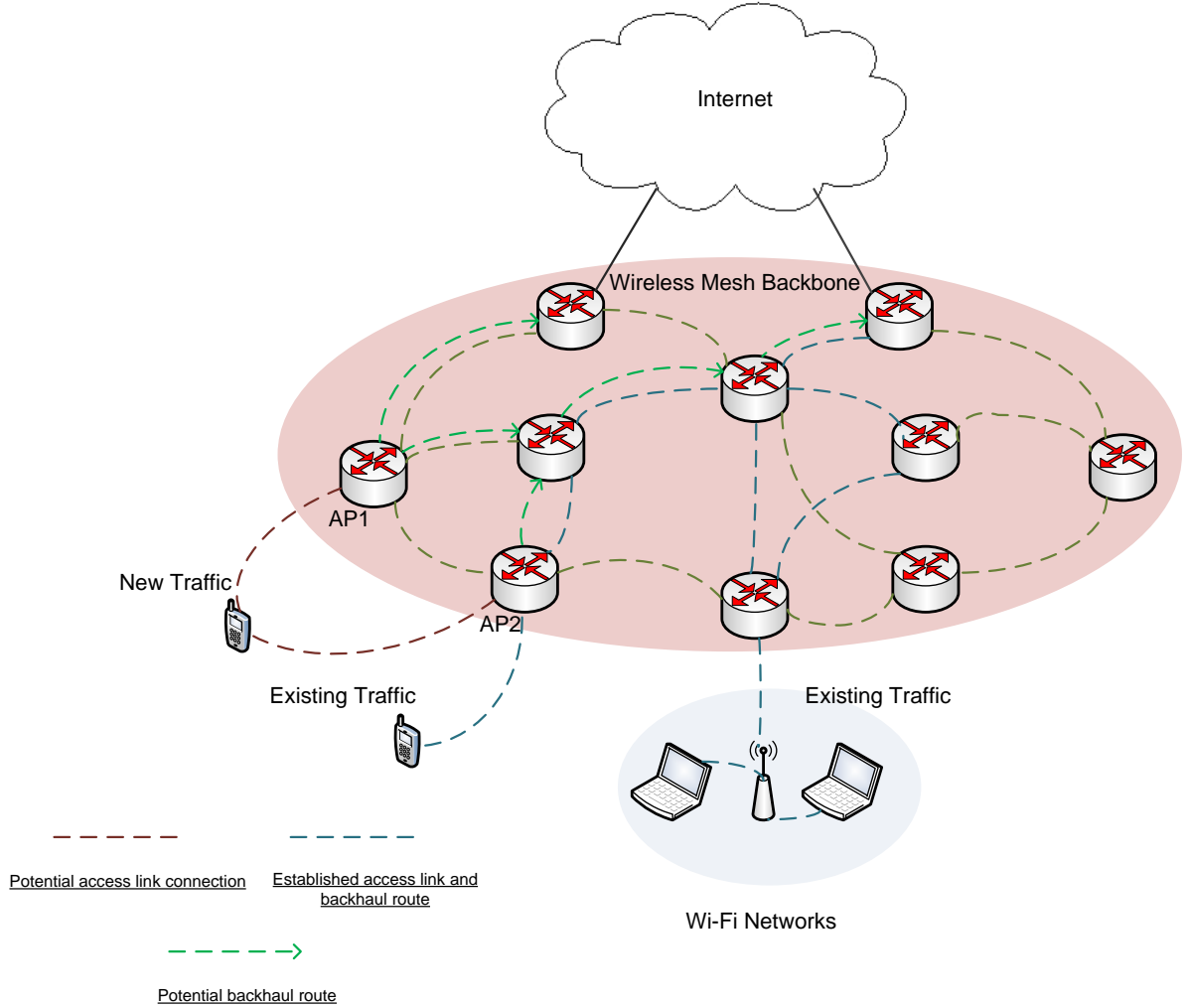


Figure 2.7: CAC and routing in WMNs

in both the access links of AP_1 and AP_2 as well as the available capacity in backhaul routes should be made. In Figure 2.7, there are two potential APs that are within the transmission range of the new traffic and three potential backhaul routes. Depending on how much bandwidth is being used by the existing traffic in the backhaul and how much residual capacity is available as a result on each of the three potential backhaul routes, the new traffic is either rejected or admitted into the network via one of the APs. As mentioned earlier in Figure 2.7 routes

were set up based on a specific routing metric before the arrival of the new traffic and the challenge is to find the best AP that can accommodate the new traffic in both its access links and backbone route.

In this section an overview of routing in WMNs is stated. Routing can be defined as the process of configuring a path between a source node and a destination node. Challenges involved in designing a routing mechanism for different wireless networks vary depending on their characteristics. In WMNs, facilitating an efficient wireless interconnection between APs and the wired backbone through intermediate nodes is a complex task, one which is not possible through incorporating conventional routing mechanism that were used in wired and wireless networks. This is due to the shared nature of wireless medium along with the static nature of nodes. In order to address the issue of communication among mesh nodes and the gateway/s in WMNs, a routing protocol combined with a routing metric is needed, which should aim at finding the best route among all possible routes between a source-destination. The basic requirement of designing an efficient routing metric in WMNs depends on grasping two fundamental factors [36]:

1. Routing Protocols used in the wireless mesh networks: Employing different routing protocols may result in various costs such as management complexity and message overhead; therefore when designing a routing metric all the costs associated with using a specific routing protocol should be taken into account.
2. The characteristics of a mesh network from a routing perspective: As mentioned previously, the multi-hop wireless nature of WMNs imposes added complexity in designing the routing metric and demands for a different approach. The network topology of WMNs differs from other wireless technologies; wireless backbone is fixed and node mobility in the backbone is not frequent [17]. Also there is

a possibility of inter-channel interference, especially when using omni-directional antennas. For instance, a wireless link may not have dedicated bandwidth due to neighbouring nodes' transmissions, this is a well-known problem of hidden and exposed terminals. Capacity of links in any wireless communication can vary in time due to surrounding interference, in WMNs this problem is even more prominent as multiple technologies may use the same frequency band. Moreover, introducing channel diversity in routing process results in less inter-nodes interference and increased overall throughput.

In the next subsections firstly different routing protocol for WMNs are explained, then existing routing metrics for WMNs are presented in details. The main routing algorithm utilised in the rest of this study is also outlined. It must be emphasised that designing a new routing protocol or metric is out of the scope of this thesis, and it is not part of contributions made. The background related to routing in WMNs is mentioned to shed some light on the route setting up used in the following chapters.

2.5.1.1 Routing Protocols for Mesh Networks

Routing protocols are divided into three different categories depending on when they are calculated: on-demand routing, proactive routing and hybrid routing[36] [17]. Based on whether the network is organised, flat or hierarchical, a two further categories can be mentioned [17]. Below these categories are further explained.

Reactive Routing:

On-demand or reactive routing protocol was initially proposed for ad-hoc networks and is based on finding a route between a source-destination only when

the source node needs to send packets to the destination node. By incorporating flooding techniques a route can be discovered when needed. On-demand routing is best suited for ad-hoc networks as the frequency of link breaks within the network is high, which is due to the mobility of the nodes. Flooding can provide a high network connectivity and relatively low message overhead for ad-hoc networks. In contrast to the ad-hoc networks, on-demand routing based on flooding techniques is not scalable for WMNs as it can be both redundant and costly in term of message overhead. This is due to the static nature of the nodes and the low frequency of link breaks comparing to frequency of flow arrivals.

Proactive Routing:

Proactive routing protocol is a protocol where at each node one or more routing tables are saved. The routing table/s contain routing information on how to reach any node in the network and is updated periodically in order to maintain an updated perspective of the connections in the network. Based on the method of packet forwarding along routes this routing protocol can be further divided into two subcategories: I. Source Routing [37] II. Hop-by-hop Routing

Hybrid Routing:

This protocol is a mixture of the above two mentioned protocols. In hybrid routing some of the nodes may use proactive routing and some may adapt on-demand routing.

Flat Routing:

In flat routing all the nodes in the network are on the same level and have the same routing roles. This is a really simple protocol but since all nodes maintain global routing information, it may not be scalable.

Hierarchical Routing:

Some nodes or a cluster of nodes maintain global routing information but the local routing information is kept in all the nodes. This protocol is more scalable but it suffers from design complexities.

2.5.1.2 Routing Metrics

Routing metrics are integrated in routing protocols, and determine the objective of the routing. According to [38] routing metrics in WMNs can be classified into seven different types. Below the most widely used types are summerised [17] [38]:

- **Distance Routing Metric:**

The philosophy behind this routing metric is based on finding the path with minimum distance between a source-destination. Hop Count metric has been employed in different routing protocols to find the minimum distance between a source-destination. In Chapter 3 and 4, shortest path routing based on Dijkstra's algorithm [39] (Chapter 5, Section 5.2.3), is implemented. Dijkstra's algorithm is explained in details in the next subsection.

- **Latency Routing Metric:**

As the name suggest this class is based on minimising the latency encountered when routing packets. Two different latency routing metrics exist in WMNs, Per-hop Round Trip Time (RTT) [40] and Per-hop Packet Pair Delay (PktPair) [41].

- **Error Rate Routing Metric:**

This metric addresses the issue of error in routing packets in wireless com-

munication. A well know metric of this class is Expected Transmission Count (ETX) [42], which estimates the expected number of MAC re-transmissions a node needs in order to successfully transmit a packet to a destination node. Expected Transmission Time (ETT) [43] is the enhanced version of ETX and characterizes the expected MAC transmission time of a packet of a known size over a link.

- Multi-channel Routing Metric:

None of the mentioned metrics capture the effect of interference. Weighted Cumulative Expected Transmission Time (WCETT) was proposed by [43] and counted for the multi-channel nature of the WMNs. It basically changed ETT to capture intra-flow interference.

2.5.1.3 Dijkstra's algorithm

Dijkstra's algorithm is utilised in finding the shortest path between each source-destination in the rest of this thesis. In this section the background related to this algorithm is stated.

Dijkstra's algorithm finds the shortest paths in a graph with non-negative path length [39](Chapter 5, Section 5.2.3). The main objective is to find the shortest path from a source node to all the nodes in a graph. The algorithm starts off by finding the shortest of the shortest paths to the source node which is a single arc-path from the source node. Having found the first closest node to the source node, the next shortest path is either a single-arc path from the source node or a shortest two-arc path form the previously chosen node. The search continue with the same pattern until all the shortest paths are found. The procedure is as follows:

- Each node $_i$ is assigned with an estimate shortest path length D_i to the source node S .
- A set P is defined which demonstrates the permanently labeled nodes. Initially, the only permanently labeled node is the source node (i.e. $P = \{S\}$)
- At each stage the nodes that are added to P are the closest nodes to S that are not already in P .

The algorithm works as follows:

- Initially $P = \{S\}$, $D_s = 0$ and $D_j = d_{js}$ for $j \neq 1$
- Step 1: Find $i \notin P$ such that $D_i = \min_{j \notin P} D_j$
Set $P := P \cup \{i\}$, If P contains all the nodes then the algorithm is complete and the search must be stopped.
- Step 2: For all $j \notin P$ set
 $D_j := \min[D_j, d_{ji} + D_i]$
Go to Step 1.

A new node is added to the P at each iteration of the algorithm; therefore assuming that there are N nodes in the graph the algorithm terminates after $N - 1$ iterations. The computational complexity of the Dijkstra's algorithm is $O(n^2)$, given that there are $N - 1$ iterations and the number of operations per iteration is nearly N .

2.6 CAC in all-IP Networks with Mobility Support

As stated earlier in this chapter, in the first two generations of wireless systems CAC schemes were developed for voice services. CAC schemes designed for 3G networks were based on providing QoS for multiple-class services such as video, audio, voice and data. Future IP based mobile networks consist of various wireless technologies that must operate with each other to provide an efficient seamless connectivity. One of the main challenges of designing a CAC scheme in all-IP networks is driven by heterogeneity, providing QoS requirement for multimedia applications at the presence of different wireless access technologies is a difficult task. Also future networks inherit IP's lack of QoS support. Two main architectures for QoS provisioning in Internet are Integrated Service (IntServ) [44] and Differentiated Service (DiffServ) [45]. The IntServ provide strict QoS guarantees per flow, which involves per flow CAC signalling and resource reservation at intermediate nodes through reservation protocols such as Resource Reservation Protocol (RSVP) protocol [46]. The per flow based nature of IntServ made them unscalable. DiffServ operates on group of flows as oppose to per flow, and it aggregates IP flows with similar QoS requirements into the same group. The classification of IP flows is performed on the network edge routers and less complex operations are handled by the core routers. However, these two techniques were designed for static networks and are not readily applicable to the wireless networks [47]. This impose extra challenges in QoS provisioning in future all-IP networks and therefore different mobility aware IntServ and DiffServ solutions were proposed to extend these architectures to accommodate mobility.

Challenges involved in designing an efficient admission control scheme in all-IP

2.6.CAC in all-IP Networks with Mobility Support

networks with mobility support depends on the specific protocol and the class or family of mobility management solutions. The available bandwidth in such networks is affected by signalling over the wireless link and by the bottleneck effect caused by presence of MAs. As previously stated in section 2.4, in Mobile IPv6 the main issue in providing an efficient admission control is reducing the high probability of handover failure as well as the handover latency due to frequent signalling between the MN and HA. This led to the birth of micro mobility family of protocols, and MA based solutions became the leading solution in micro mobility protocols. Pragad in [48] showed that implementing MA based micro mobility solutions results in the reduction of the capacity of the wireless access network. Instead of routing the packets through the best routes, packets are routed through the MA/s and as a result the level of congestion in the network increases. The presence of the MAs in PMIPv6 networks also creates the bottleneck effect and reduces the overall capacity of the network by overutilising certain nodes and paths and leaving the rest of network underutilised. Therefore, when designing a CAC scheme for MA based micro mobility solutions the effect of employing MAs must be taken into consideration. An efficient CAC mechanism aims at eradicating the bottleneck effect by either reducing congestion level at the local MA through using appropriate resource allocation techniques or by directly binding MNs that do not require stringent mobility support with the CN. RO techniques as suggested in [49] [50] [51] and [52], can be incorporated to facilitate more resources for admission.

Chapter 3

Joint Revenue-based Call Admission Control and Routing

3.1 Introduction

QoS provisioning mechanisms play an important role in ensuring the proportional rationing of resources amongst multiple users of different types of traffic according to their requirements. One of the fundamental mechanisms of QoS provisioning is CAC which restricts the access to the network based on the availability of resources [7]. WMNs are expected to support variety of traffic types with different QoS requirements [16]. Hence, having an efficient admission control scheme that can provide a guaranteed QoS is paramount. Designing CAC algorithms for WMNs is a challenging task given the scarcity of resources in both the access and backhaul links and the shared nature of intermediate backhaul nodes and wireless medium. To this end, CAC and routing in WMNs are strongly interrelated and a joint design scheme can strive to maximise the performance of the system.

In this chapter, a joint CAC and routing algorithm design is considered, where routing paths are pre-constructed in a separate phase and CAC is formulated as an optimisation problem.

Next generation of wireless networks would have to offer a vast range of multimedia services to the users; hence it is vital to design a CAC algorithm that besides satisfying the users' QoS requirements, would aim to maximise revenues (monetary incentives) of the network operators. Under this perspective, this research study focuses on defining a maximisation problem, where the total revenue from all the carried connections in the network need to be maximised whilst taking into account the bandwidth constraints imposed by the wireless access and the backhaul links. Majority of the research in the context of CAC consider maximising the total revenue gain, controlling congestion level and designing pricing policies separately [53]. The focus is either on tuning the amount of traffic that is admitted in the network and the resulting revenue gain or on designing different pricing policy models. Little work has been done where both factors and their impact on each other were taken into consideration. In this study there is also an emphasis on the effect of different pricing models on the performance of the system, such as connection blocking probability, and investigating the economical side of users' management in CAC.

The investigated scenario in this chapter assumed a batch based CAC approach and addressed the following issues:

1. The joint admission control and route assignment is formulated as a Multiple Knapsack with Assignment Restrictions (MKAR) problem [54]. Firstly, K shortest paths for each source-destination pair are constructed. An optimisation problem is then detailed based on the MKAR problem which attempts to find the maximum total revenue gain by the network provider whilst satisfying the

capacity constraints in the access and backhaul links.

2. Effects of different pricing models on the performance of the system were investigated by using three main classes of randomly generated instances for 0-1 Knapsack Problems and employing it on a single Mesh Access Point (MAP). Through simulation studies it has been shown that the uncorrelated random class of weighing significantly outperforms the other two instances in both the total revenue gain and blocking probability, and therefore this pricing policy is chosen to be the optimal route.
3. Using the pre-computed paths the optimal solution through exact method is achieved.
4. To provide a solution for the proposed optimisation problem in the real time, a heuristic approach called Greedy Call Packing (GCP) [55](Chapter 2, Section 2.4) is used at first in order to find an initial solution. Pursuing a more feasible solution, the initial solution achieved by the GCP is incorporated by a meta-heuristic approach called Simulated Annealing (SA) [56].
5. Through extensive simulation studies it has been shown that batch processing can considerably improve the system's performance and incorporating meta-heuristic approaches results in a better near-optimal solution.

The remainder of this chapter is organised as follows. Section 3.2 provides the motivation for this work and a summary of relevant works in the context of revenue-based CAC and pricing in telecommunication networks. This is followed by a formal definition of the problem in section 3.3. The system model is detailed in section 3.4, which includes modelling joint CAC and route assignment as an MKAR. The routing algorithm implemented alongside the pricing models investigated are included in this section. Section 3.5 sets out the framework used to evaluate the performance of the proposed scheme. The results of this evaluation are also presented in this section and finally this chapter is concluded in section

3.6.

3.2 Motivation and Background

In this section the relevant work concerning the contributions of this chapter is outlined. One of the main objectives of network providers is to maximise their revenue. This can be achieved either through CAC or pricing. In this respect, the literature behind revenue-based CAC is presented in section 3.2.1. The total revenue gained by the network provider is inherently related to the pricing strategy adapted by the service/network provider. Section 3.2.2 details the issue of pricing in the telecommunication networks.

3.2.1 Revenue-based CAC

Admitting a new call into the network has associated rewards and penalties from the network provider's perspective. The reward can be driven from utilisation of network resources for the purpose of revenue generation. Under certain network conditions when a new call is admitted into the network, it may result in deterioration of QoS of existing calls as well as potentially raising the congestion level in the network. As a result dropping probability or future blocking probability may increase, therefore, there is a potential for associating penalties for admitting calls under specific network conditions. It must be highlighted that both "reward" and "penalty" are generic terms, which refer to a certain value or loss brought to the system respectively. In this class of CAC, based on the potential rewards and/or penalties the network revenue can be maximised. The admission criterion depends on the network condition and the manner in which the problem

is defined, for instance it can be the number of users [5].

CAC policies can have a major effect on the revenue gained by the network providers. Different approaches to revenue-based CAC were proposed in the literature. Some of these approaches are either based on minimising the penalty associated with rejecting a call or a class of calls, some are based on maximising the reward associated with admitting a call or a certain class of calls and in some a joint reward/penalty is defined. For instance, in [57] revenue maximisation is achieved through minimising the penalty of dropping an existing call and channel reassignment. A comparative analysis of reward-based CAC designed for servicing multiple priority classes in wireless networks with QoS guarantees is carried out in [58]. The authors state that reward is a generic term and emphasise that in their work reward is defined as the revenue earned by the service provider and is modelled through the “charge-by-time” scheme. Moreover, performance of partitioning, threshold-based, spillover, and elastic threshold CAC algorithms in terms of revenue (i.e. reward) maximisation is analysed. In [59], a cost-based CAC scheme based on rewards-penalties associated with various classes of multimedia service is proposed. The objective of their research study is to maximise the total reward in the network, and in this respect resources are reserved for different media type requests based on their predefined priority which is value/penalty.

One of the well studied approaches to revenue-based CAC is incorporating prioritisation schemes, especially with different classes of service being present in the network. Under this perspective, a prioritisation scheme is designed which favours the revenue gain of the service provider. For instance, in [60] a batch and prioritise scheme is proposed where revenue is maximised. In this approach the adaptive priority of a class of traffic depends on the tolerance degree of that class to delay as well as link-utilisation. Another approach to revenue-optimisation through

CAC is by means of channel reassignment. A revenue-based CAC for wireless cellular networks, using channel reassignment is proposed in [61], where the study demonstrates the relationship between the number of guard channels, reassignments and total revenue. Reservation-based schemes have also been widely employed in research studies in the context of revenue optimisation. An end-to-end bandwidth reservation CAC is introduced in [62], where one of the objectives is to maximise the revenue gain. In their proposed scheme, different fractions of bandwidth are exclusively reserved for different classes of traffic; this is based on service differentiation defined in the IEEE 802.16 networks.

Integrating pricing with admission control is an efficient approach in addressing revenue-based CAC, given that call request arrivals to the system can be controlled through monetary incentives. The integrated CAC and dynamic pricing in [63], aims to maximise the revenue and reduce the congestion in wireless cellular networks. The stochastic dynamic formulation proposed, considers the effects of pricing on call arrivals, retrials and substitutions among services and time. The behaviour of users from different classes towards prices is modelled by their Willingness to Pay (WTP). Overall their strategy provides a platform where low-WTP users as well as high-WTP users can access the network depending on the load in the network. The CAC algorithm integrated with pricing for revenue optimisation in [64], assumes that a “charge-by-time” pricing scheme is being used by the service provider and solves the problem by partitioning and threshold-based admission control schemes. However, none of these approaches have jointly looked at revenue-based CAC and the computational complexity that various pricing strategies impose on the system as well as blocking probability resulted from employing different pricing strategy.

3.2.2 Pricing in Telecommunications networks

More importance has been recently attached to the pricing of different classes of services offered by the network provider, as resources are limited and total demand tends to be higher than the available resources. The bandwidth allocated to a user can be viewed as a commodity sold to the user by the network provider [65]. What makes the pricing problem in communication networks a complex task is the need for considering several issues at the same time; such as engineering issues, financial issues, a diverse set of desired services (i.e. elastic versus real-time traffic), and other issues such as design simplicity [66]. Pricing is an effective way of controlling the flow into the network and managing congestion. The irritation in not receiving the highly priced services is great from the users' point of view. Users' reaction to the tariff proposed by the provider indicates that increased prices would result in fewer number of users, whilst also low-priced tariffs lead to low revenue, which would consequently affect the load on the network [67].

Majority of the research in the context of pricing aims at either controlling the rate of new connection requests or regulating the data rate sent by existing connections [68]. Pricing strategies can be divided into two general categories: Static or fixed and Dynamic, this is based on the time scale over which the pricing strategy is developed. The major advantage of the static pricing strategy is its simplicity, however prices are independent of the current state in the network and this type of pricing is normally criticized for not providing enough incentives for users to avoid congestion. Dynamic pricing is a well studied area and several methods have been proposed in the literature addressing this category of pricing; such as agent based, time-of-day, usage-based, threshold-based and congestion dependent pricing [68]. This category of pricing is often criticized for being too complex and for causing user dissatisfaction as a result of frequent changes in the price of the service [69].

Both these pricing categories mentioned have been extensively studied in the literature; however the issue of pricing strategy and the computational complexity it can impose on the system has not been addressed previously. The correlation between the amount of bandwidth and the price is a good indicator for the level of computational complexity that certain mathematical models can potentially have. This was taken into consideration in designing an optimal pricing policy in this study and is further explained in 3.4.3

3.3 Problem Description

The knowledge of available bandwidth and congested links in the network is beneficial in many ways. For instance, the admission decision of a new flow into the network depends on the available bandwidth. Also, the decision as to which path a flow should be routed through depends on the available bandwidth and the level of congestion on different links. Bottleneck link is the most congested link along a path or in a network, which can limit the amount of flow that can be routed/admitted through a path or into the network. The presence and location of bottlenecks in the network depends on the design criteria of the network as well as the characteristics and nature of it. In a single-hop WLAN, APs are directly connected to the wired backbone and the access link is most likely to be the bottleneck. In WMNs however, the traffic can be bottlenecked by either the access or the backhaul links. Measuring the amount of available bandwidth therefore requires a combined knowledge of available bandwidth in the access and the backhaul links. Joint association of each mesh client with a MAP and multi-hop backhaul routing to the Gateway, determine the availability of resources as to admit or reject a flow. Thus, the central theme of this chapter is to design a

joint admission control and routing that avoids the bottlenecks.

The multi-hop nature of WMNs enables using multiple paths between a source-destination which is paramount since using a single path between a source-destination fundamentally reduces the throughput and reliability of the network [70]. With the aid of multipath routing, network resources can be utilised and the overall blocking probability of the network can be reduced. Load balancing and providing QoS for different classes of traffic are other benefits of multipath routing. In this chapter with the aim of increasing admission resources and reliability of WMNs networks, a multipath routing algorithm is implemented. Details of this routing algorithm are mentioned in the next section.

Admission control decision needs to be made with a certain objective. A wide range of multimedia services are being supported by the next generation of wireless networks and the economical incentives of network providers are increasing. Hence, it is vital to design a CAC algorithm which besides satisfying the bandwidth constraints aims at addressing the monetary incentives of the network providers. Two approaches can be used to address the cost benefits to the network provider: 1) Designing a revenue-based CAC algorithm 2) Investigating different pricing policies. The proposed model in this chapter is based on the idea that if the cost benefits of the network provider is the topic in mind, these two approaches should be jointly considered. Pricing can change the rate of flow into the network, for instance though low priced tariffs may result in high revenue, they can also raise the congestion level in the network and result in high blocking probability. Under this perspective, one of the aims of this chapter is to design a pricing policy that strives to maximise the total revenue of the network provider whilst not resulting in a high blocking probability. Once this pricing policy is found, the CAC algorithm with the objective of revenue maximisation

is implemented.

3.4 System Model

This section gives a formal definition of the models used for performance evaluation study. In this study the joint admission control and route assignment problem is formulated as a Multiple Knapsack with Assignment Restrictions (MKAR) problem [54], which is a variant of the infamous Multiple Knapsack Problem (MKP) as explained in Chapter 6 of the book by Martello and Toth [55]. The main difference between MKAR and MKP is that in MKP any item can be assigned to any knapsack. In MKAR, a set of items $N = [1, 2, \dots, n]$ and a set of knapsacks $M = [1, 2, \dots, m]$ with positive real weight w_i and capacity c_j are given respectively. For each item $i \in N$ a set of knapsacks $A_i \subseteq M$ that can hold item i is specified. Item i is known to be admissible to knapsack j if $i \in B_j$, where B_j is the set of items that can be assigned to knapsack j denoted by $B_j \subseteq N$. The goal is to choose a subset S_j of items that can be assigned to each knapsack j , with the objective of maximising the assigned weight such that three constraints are satisfied: 1) Each item is assigned to at most one knapsack (i.e. all S_j 's are disjoint). 2) Each item is assigned to one of the knapsacks it is admissible to. 3) Capacity restrictions of knapsacks are not violated, which means that the total weight of the items assigned to a knapsack does not exceed the knapsack's capacity. In the model proposed in this chapter, MAPs are considered as the knapsacks with a limited capacity and flows as the items that need to be fitted in. The price associated with each flow is considered as the profit, and the bandwidth required by each flow is the weight of each item. Contrary to the model proposed in [54], the profit associated with assigning an item is not

assumed to be equal to the weight of the item and therefore the objective is to maximise the profit of the network provider. Given that not all the flows are within the transmission range of all MAPs, a set of flows per MAP is defined, which is the same as a set of items being admissible to a specific knapsack or set of knapsacks.

Upon arrival of a request (i.e. a flow), a decision as to which MAP to admit a flow and which shortest path to route a flow through should be made. This process is based on the available bandwidth in both the access link and the backhaul links. Jointly solving the admission control and routing in a wireless multi-hop environment, given a set of nodes and demands, is an NP-hard optimisation problem [35]. Therefore, we decompose the problem into routing (path construction) and admission control, thus solving them separately. Once the set of paths are constructed for each source-destination node, the problem becomes a joint admission control and route assignment problem. The goal is to associate each flow with an appropriate MAP which has enough bandwidth in both its access link and routing path from the gateway to accommodate that flow, in order to maximise the revenue from all the carried connections.

As mentioned previously pricing is a method of controlling the flow rate into network and it can have several non-monetary impacts on the system such as high blocking probability or a highly congested network. These impacts depend on the pricing model used as well as the model that the CAC algorithm is based on. Since the system in this chapter is modelled as an MKAR, the correlation between the price and the bandwidth (rate) of each flow also needs to be considered, as this will change the hardness and complexity of the problem [71]. The prime goal is to find a pricing policy that maximises the total revenue gain but which also takes into consideration the connection blocking probability and the hardness of

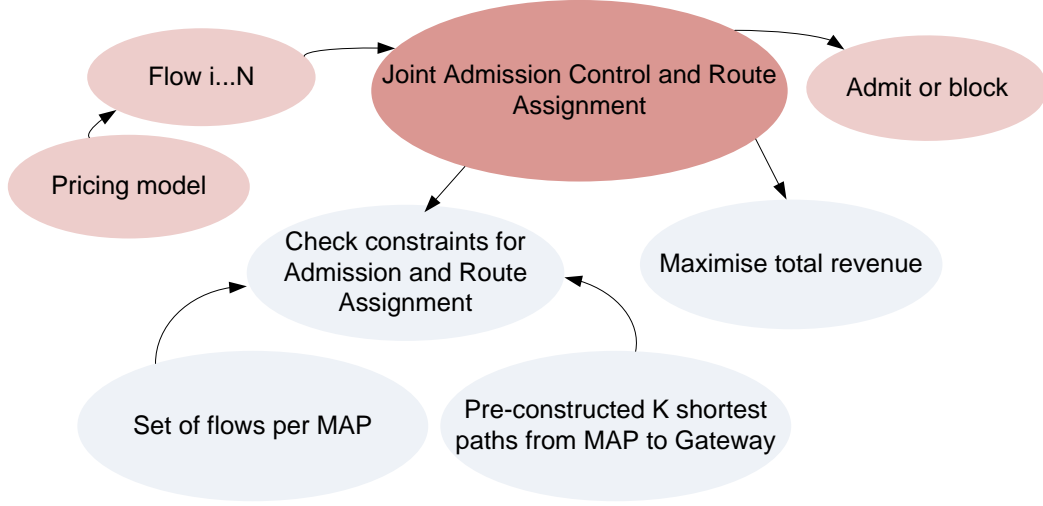


Figure 3.1: Overview of joint Admission Control and Route Assignment Decision

the problem. An overview of the above mentioned models is illustrated in 3.1, followed by a detailed explanation of the models in the next subsections.

3.4.1 Network Model

The wireless mesh network is represented as a connectivity graph $G(V; L)$, where V is a set of vertices (wireless nodes) and L is the set of edges or transmission links that satisfy a pre-defined SINR threshold criterion. There is a set of $m = [1, 2, \dots, M]$ MAPs and $n = [1, 2, \dots, N]$ flows (or mesh clients). One of the fundamental assumption is that each one of the wireless nodes use the same modulation scheme and transmit with the same fixed transmission power; thus a further assumption is that each node has an identical fixed transmission range of r_{tx} , which is a well-used assumption in wireless multi-hop networks [72][32].

A logical link between node u and v exists if the distance between them is less than or equal the r_{tx} :

$$\begin{aligned}
 L &= \{(u; v) | u, v \in V\} \\
 \text{s.t. } & u \text{ being in } r_{tx} \text{ of } v \text{ and vice versa.}
 \end{aligned} \tag{3.1}$$

The assumption of slow fading channel on the backhaul links is made, which is a realistic assumption as the end points are static. Let C_j and $b(v)$ be the capacity of each MAP access link and each wireless node in the backbone respectively, where $j \in m = [1, 2, \dots, M]$. The bandwidth request of each flow $i \in n = [1, 2, \dots, N]$ is denoted by w_i and the price associated with admitting the corresponding flow is denoted by p_i . Using the routing algorithm, $k = [1, 2, \dots, K]$ shortest paths (sp) from each origin to the destination are constructed.

In order to be able to define the problem in a mathematical programming setting, the following Boolean variables are introduced, in which sp denotes the calculated shortest path:

$$D_{ij} = \begin{cases} 1 & \text{if flow } i \text{ is within transmission range of MAP}_j \\ 0 & \text{Otherwise} \end{cases} \tag{3.2}$$

$$sp^v(jk) = \begin{cases} 1 & \text{if } sp_k \text{ of MAP}_j \text{ passes through node } v \\ 0 & \text{Otherwise} \end{cases} \tag{3.3}$$

$$x_{ijk} = \begin{cases} 1 & \text{if flow } i \text{ is assigned to } sp_k \text{ of MAP}_j \\ 0 & \text{Otherwise} \end{cases} \tag{3.4}$$

In equation 3.2 D_{ij} is the transmission range Boolean variable, that shows which flows are within the transmission range of which MAPs. A sp_k of a MAP_j might share an intermediate node with another sp of the same or a different MAP and becomes a bottleneck. As stated before routing the flows over one of the k sp s is

one of the the main interests; therefore 3.3 and 3.4 exhibits whether a sp_k passes through node $v \in V$ and the decision variable x_{ijk} respectively. Each flow i can only be admitted to one sp_k of a MAP_j for which D_{ij} is equal to 1. The decision variable x_{ijk} demonstrates which sp_k of a MAP_j , flow i is assigned to.

Batch processing results in a significant degree of improvement regarding the admission control and revenue maximising as opposed to handling the flows individually. This was investigated by means of extensive simulation studies and results where compared to an on-line scheme as shown in section 3.5.2. In the on-line scheme, the decision as to reject or admit a request is made at the instant of its arrival. This decision is independent from other prior or future arrivals and is based on the available resources as well as the state of the network. It is further assumed that flows arrive according to a Poisson process with arrival rate λ , and the optimisation problem is solved for each batch. Within a batch, flows are admitted with the objective of revenue maximisation and are dealt with in any order and not according to the arrival order.

3.4.2 Routing Algorithm

As mentioned earlier, joint consideration of routing and CAC is an NP-hard problem. It involves taking the wireless interference constraints into account for configuring and constructing the paths on demand as well as making an admission control decision for achieving the desired objectives. In the proposed model routing paths are pre-constructed in a separate phase, the admission of demands is then carried out assuming that our environment is interference free. The assumption of an interference free environment and pre-computed paths [35] for each source-destination has also been adopted in [34].

Multipath routing enables using several sub-optimal paths to reach a destination as oppose to using one best or optimal path [70]. In this chapter in order to increase the network reliability and available resources a multipath routing algorithm is implemented. The objective of the implemented multipath routing algorithm is to find K shortest paths between each MAP-Gateway. The problem of determining K shortest paths between a pair of nodes in a network is a problem where for a given $K \geq 1$ the intention is to determine successively the shortest path, the second shortest path, \dots , until the K -th shortest path between the given pair of nodes. The cost metric used here is the free path loss for the link l . The algorithm in this chapter, manages to find K shortest paths for each source-destination pair based on a modified version of Yen's algorithm [73] and similar to the one proposed in [74]. Dijkstra's algorithm [39] (Chapter 5, Section 5.2.3), manages to find the shortest path tree from a source node to every other node and therefore Dijkstra's algorithm is utilised in finding the shortest path between each source-destination.

The K shortest paths constructed are not node-disjoint and they might share an intermediate node with any other shortest path in the topology. This phenomena imposes further constraints on the available resources of the K shortest paths and therefore the basic assumption of all paths having a certain and equal available bandwidth can no longer be valid. One of the primary focuses of this work is to identify the bottlenecks in the system and in this regard a matrix of all $sp - v$ is constructed for the specific WMNs topology, which is expressed by the Boolean variable 3.3. The $sp - v$ matrix is a 0-1 matrix that shows the nodes each shortest path in the network has to pass through. With the aid of this matrix the true available bandwidth along a shortest path can be identified and the decision as to which path to route a flow through can be performed efficiently.

3.4.3 Price and Revenue Function

Pricing for different telecommunication services is a topic that has been widely studied in the literature and influences how QoS and CAC are intertwined [53]. It can be viewed as a way of monitoring the amount of flows into the network. This will have a direct impact on congestion levels and the total throughput of the network.

A similar pricing demand-function to the one proposed in [75] is used here, which is detailed as follows:

$$\lambda_\beta = a_\beta (p_\beta)^{-\varepsilon_\beta}, \quad (3.5)$$

where λ_β denotes the arrival rate of class β traffic and p_β denotes the price associated with class β traffic. Also, a_β and ε_β are the constants correlating the variables λ_β and p_β . From equation 3.5, it can be immediately gathered that pricing a class of traffic determines the arrival rate of that class of service. The lower the price associated with a certain class of traffic, the higher the arrival rate of that certain class of traffic. This approach can result in a higher revenue gain but also low pricing can result in a highly congested network, with a high blocking probability [76]. In this work a static pricing proportional to the amount of bandwidth that a user requests is considered, as dynamic pricing might cause user dissatisfaction due to frequent changes in the price of the service [69].

In the proposed model here, p_i is the price associated with admitting flow i . The total revenue that the operator gains is defined as follows, which is the main objective of the CAC algorithm in 3.4.4:

$$\text{Maximise } \sum_{j \in \Omega_i} \sum_{k=1}^K \sum_{i=1}^N x_{ijk} p_i$$

where:

$$\Omega_i = \{\text{MAP}_j | D_{ij} = 1\}$$

As outlined before the WMNs is modelled as an MKAR problem, thus the correlation between the profit and weight also needs to be considered in this type of optimisation problem and the complexity that this would impose on different instances of the problem. The computational experiments proposed in the literature in the context of 0-1 Knapsack Problems commonly consider three main classes of randomly generated instances [71]:

1. Uncorrelated :

w_i uniformly random in $[b, a]$

p_i uniformly random in $[b, a]$

$$C = \frac{1}{2} \sum_{i=1}^n w_i$$

2. Weakly Correlated:

w_i uniformly random in $[b, a]$

p_i uniformly random in $[w_i - \delta, w_i + \delta]$

$$C = \frac{1}{2} \sum_{i=1}^n w_i$$

3. Strongly Correlated:

w_i uniformly random in $[b, a]$

$$p_i = w_i + \delta$$

$$C = \frac{1}{2} \sum_{i=1}^n w_i$$

The strength of correlation between the weight and profit of an item in 0-1 Knapsack Problems is an immensely important factor. Stronger correlation between the profit and weight imposes more complexity and hardness in finding the optimal solution; therefore three factors should be considered in finding an optimal

pricing policy: the total revenue of the operator, the blocking probability in the network and the complexity of the problem. In order to investigate this further, first the 0-1 Knapsack Problem was solved for the three instance using varied number of flows up to 10000. The results showed that in the strongly correlated instance for more than 100 flows the algorithm does not converge and finding an optimal solution in real time is not possible. Therefore, a maximum of 100 flows are considered and the 0-1 Knapsack Problem is solved for the three instances mentioned above. A pricing policy that maximises the revenue and at the same time does not result in a high blocking probability is then chosen based on the simulation results. It is worth mentioning that even for the 0-1 Knapsack Problem solved in this chapter, the bandwidth required by each flow and the price associated with it are considered to be the weight and profit of each item respectively.

3.4.4 CAC Algorithm

In this section the joint route assignment and CAC algorithm is formulated as follows:

$$\text{Maximise } \sum_{i=1}^N \sum_{j \in \Omega_i} \sum_{k=1}^K x_{ijk} p_i \quad (3.6)$$

$$\text{Subject to } \sum_{i=1}^N \sum_{j \in \Omega_i} \sum_{k=1}^K w_i x_{ijk} \leq C_j \quad (3.7)$$

$$\sum_{i=1}^N \sum_{j \in \Omega_i} \sum_{k=1}^K sp^v(jk) w_i x_{ijk} \leq b(v), \quad \forall v \in V \quad (3.8)$$

$$\sum_{j \in \Omega_i} \sum_{k=1}^K x_{ijk} \leq 1, \quad \forall i \quad (3.9)$$

$$x_{ijk} \in \{0, 1\}, \quad \forall (i, j, k) \quad (3.10)$$

$$D_{ij} \in \{0, 1\}. \quad \forall (i, j) \quad (3.11)$$

Constraints 3.7, 3.8 and 3.9 state that a flow i can only be admitted if there is enough capacity in the access link of the MAP_j as well as the sp_k of the MAP_j . 3.10 states that a flow can only be admitted to one of the K sp s of a MAP_j . A set of MAPs per flow is defined as $\Omega_i = \{\text{MAP}_j | D_{ij} = 1\}$, which shows the number of MAPs that a flow is within the transmission range of. The first constraint, 3.7 finds the set of MAPs and their corresponding shortest paths that a flow i is admissible to, which potentially has enough capacity for admitting flow i . Constraint 3.8 further checks all the nodes across each shortest path of a potential MAP and as stated earlier i is admitted if both the access link C_j and sp_k capacity suffice.

3.4.5 Complexity of the CAC Algorithm

Dawande et al. in [54] states that even in cases like this study where all the MAPs have equal capacities, the MKAR algorithm is still NP-hard. This is proven by modelling the assignment restriction as a bipartite graph where the two disjoint node sets of the bigraph correspond to the sets of flows (N) and MAPs (M). If $G(V, L)$ is the resulting bipartite graph with $V = N \cup M$, then edge (i, j) can only exist if $i \in B_j$. The density of the bipartite graph is defined as the ratio of the number of edges in G to the number of edges in a complete bipartite. Utilising Numerical 3-Dimensional Matching, they prove that MKAR is still NP-hard even

when all the capacities of MAPs are equal and the bipartite graph which shows the assignment restriction is sparse. The CAC algorithm proposed in 3.4.4 has an added complexity comparing to the MKAR proposed by Dawande et al., in that the decision as to which shortest path to route the flow through also needs to happen parallel to the assignment of the flows to MAPs.

3.4.6 Near-optimal Solution

The CAC problem formulated in the previous section can not be optimally solved, therefore by incorporating Greedy Call Packing (GCP) algorithm and Simulated Annealing (SA) algorithm two near-optimal solutions are obtained. The GCP is a heuristic method and a variant of the well-known Greedy Algorithm (GA) [55](Chapter 2, Section 2.4) that chooses the local optimum, assuming that this is the global optimum it progresses making one greedy choice after another, the details of which can be found in Algorithm 1, in Appendix A. In the traditional GA algorithm where p and w denote profit and weight of the items, all items are sorted according to their greedy factor G , which is calculated as follows:

$$p = [p_1, p_2, \dots, p_n]$$

$$w = [w_1, w_2, \dots, w_n]$$

$$G = [p_1/w_1, p_2/w_2, \dots, p_n/w_n]$$

and in a manner to satisfy the following criteria:

$$p_1/w_1 > p_2/w_2 > \dots > p_n/w_n$$

The items are placed in the knapsack so long as the critical item S is not reached, S is the item at which the total weight in the knapsack exceeds the total capacity C of the knapsack:

$$S = \min\{J : \sum_{i=1}^J w_i > C\}$$

The only difference between GCP used here and GA is that in GCP once all flows are sorted according to the greedy factor, for each flow the MAP/s that it is within the transmission range of that flow is found. Then the flow is admitted if there is enough capacity in the access link of that MAP as well as all the nodes along one of the K shortest paths of that MAP.

In order to produce a good solution within a reasonable amount of computing time, an initial solution using the GCP algorithm is found. This is then followed by many iterations to modify this initial solution in order to find a more feasible one. Through incorporating SA algorithm, which is a local search algorithm with the advantage of not becoming trapped in the local optimum, a more feasible solution is achieved. The name and inspiration of the SA algorithm comes from annealing in metallurgy in which the solid state metal is heated and slowly cooled down until it reaches a minimum energy crystalline structure [56]. At high temperatures the particles within the metal have higher energy and are able to move around. When the temperature is carefully decreased the particles slowly arrange themselves into a highly structured lattice.

SA has been successfully applied to a number of combinatorial optimisation problems [77]. In a given combinatorial optimisation, solution X matches the current state of the solid, objective function $f(X)$ presents the energy of the current state and the control parameter T is the temperature of the solid. SA seeks to minimise the energy function through different iterations, which in combinatorial

optimisation is the objective function. At each iteration of the SA algorithm, the current solution and the new solution are compared. Better solutions are always accepted and non-improving solutions are accepted with the following probability $p = \exp(\frac{-\delta f}{T})$, which depends on the temperature parameter. By allowing worse or non-improving moves simulated annealing escapes from the local optima, but as the temperature decreased towards zero acceptance of non-improving moves decreases and the algorithm becomes similar to the greedy algorithm. The SA algorithm implemented here is based on the work proposed in [77] and is outlined in Algorithm 2, in Appendix A. At each temperature trial a randomly selected item is to be swapped with a randomly selected item that is already in the knapsack. This interchange forms the main part of the algorithm.

3.5 Performance Evaluation

To further investigate the performance of the proposed scheme in this chapter, MATLAB simulation platform is used. In the first set of simulations in order to evaluate the effect of pricing on the system performance, a single MAP is only considered. The second simulation scenario studies the joint CAC and route assignment in a wireless mesh topology built in MATLAB. In both simulation scenarios, a uniform random generator is used to generate the rate of the flows between 128 Kbps and 4.096 Mbps.

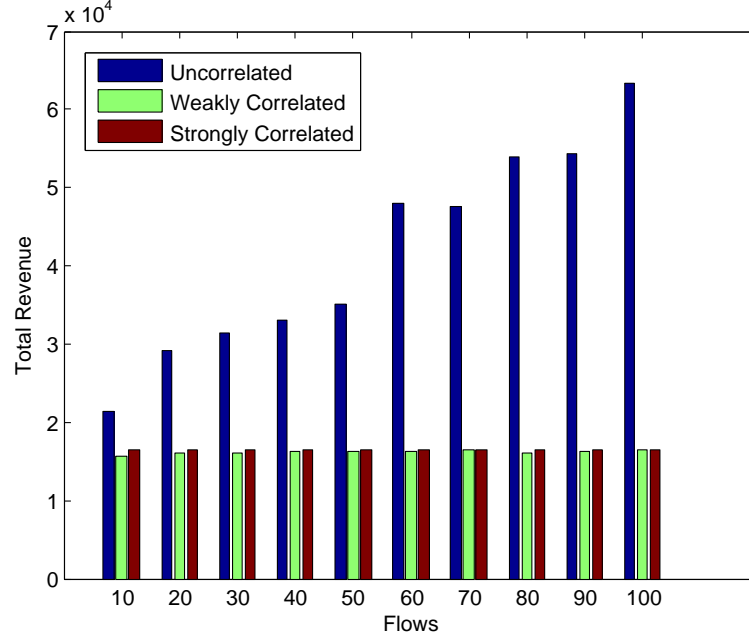


Figure 3.2: Total Revenue gained vs Flows using the three random instances

3.5.1 Simulation Scenario One: Evaluating the effect of pricing model

In the first set of simulation scenarios, CAC problem for a single MAP is modelled as a 0-1 Knapsack Problem and system performance in terms of blocking probability and total revenue is examined using 10-100 flows. The bandwidth required by each flow and the price associated with it are considered to be the weight and profit of each item respectively. Three main classes of randomly generated instances for weight and profit as discussed in subsection 3.4.3 are considered and the optimal solution using the exact method is achieved through simulations.

Results of the first simulation scenario are presented in Figure 3.2, where the effect of pricing method on the revenue is obtained. From Figure 3.2 it can be observed that on average the total revenue gained, using the uncorrelated class

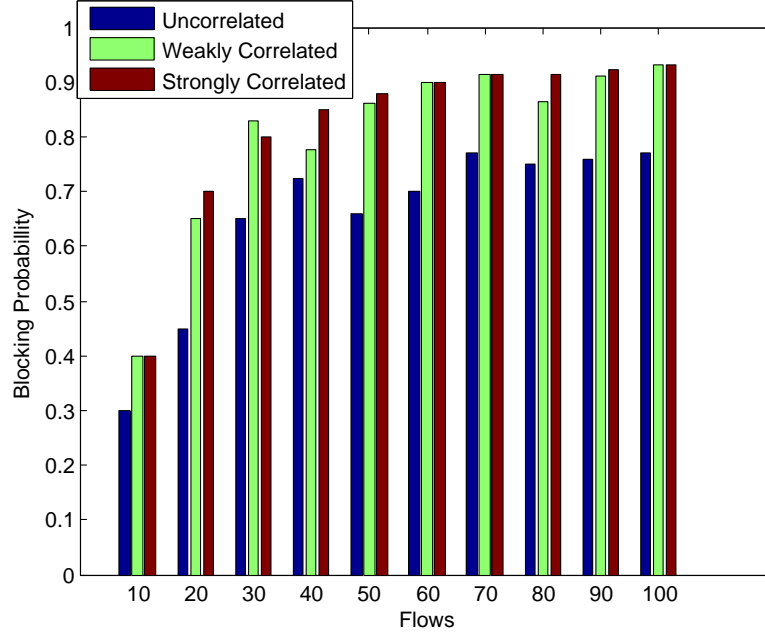


Figure 3.3: Blocking Probability vs Flows using the three random instances

of random weighings, is much higher comparing to the other two classes. This is mainly because in the uncorrelated class a wider range for price exists as opposed to the weakly and strongly correlated classes, where the price ranges $\pm\delta$ and $+\delta$ from w_i respectively.

Figure 3.3 shows how the blocking probability can be affected by the price. The blocking probability of each sample against the number of flows using the three instances is plotted. From Figure 3.2 and 3.3 it can be observed that in the scenario where there is no correlation between the price and the bandwidth the average blocking probability is the smallest. As it can be seen from Figures 3.2 and 3.3, the uncorrelated random class of weighing significantly outperforms the other two instances in both the total revenue gain and blocking probability as well as imposing less computational complexity, and therefore this pricing policy is chosen to be the optimal pricing policy and the simulation scenario two is

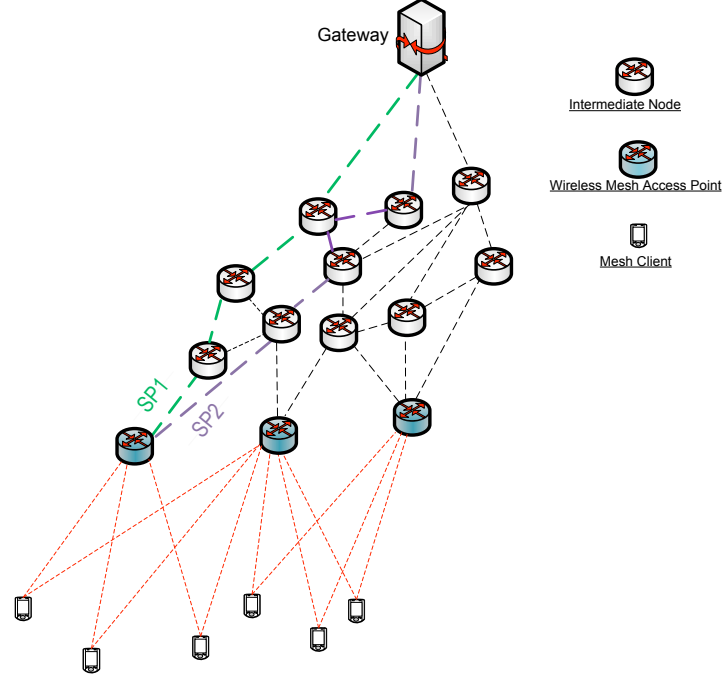


Figure 3.4: Topology of WMNs deployed in this work

based on this model. As mentioned earlier, the main objective is to maximise the total revenue that the network operator gains, and in this regard requests with higher price have more priority in terms of admission. In the strongly correlated instance, pricing is in a manner that the more bandwidth a $flow_i$ is requesting the higher the price; therefore the requests with higher price are first admitted and given that these requests require a large bandwidth they quickly use almost all of the capacity of the MAP resulting in a high blocking probability and a low total revenue gain.

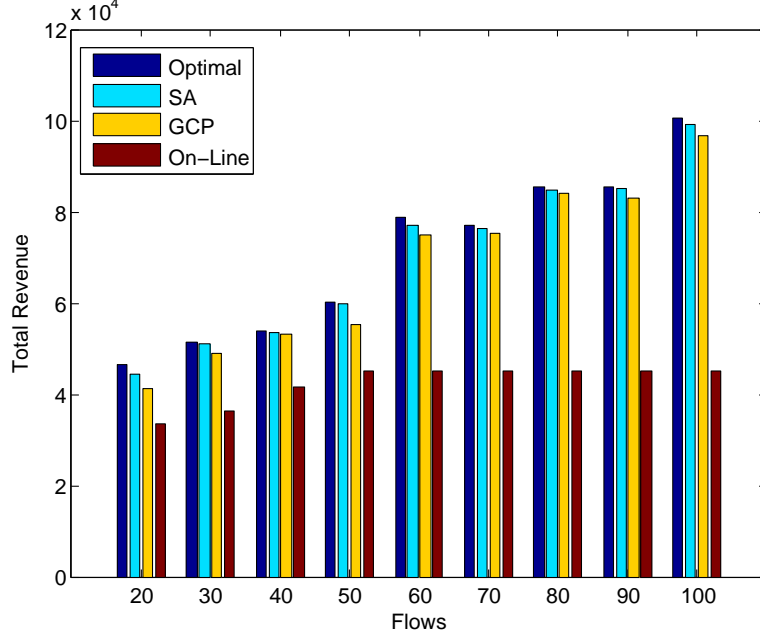


Figure 3.5: Total Revenue gained vs Flows using the optimal pricing policy

3.5.2 Simulation Scenario Two: Joint Revenue-based CAC and route assignment

The second simulation scenario consists of a wireless mesh topology with a randomly placed MAPs and randomly distributed flows in a manner that each flow is within the transmission range of at least two MAPs. The simulation environment is built in MATLAB and the resulting topology is shown in 3.4; the backhaul is a 1 km² rectangular area containing three MAPs and one gateway; there are ten intermediate nodes APs that provide the backhaul routing. Nodes use the same modulation scheme and transmit with the same fixed transmission power; therefore each node has an identical fixed transmission range of r_{tx} . Each backhaul link in the network has the equal capacity of 12 Mbps and each access link has the total capacity of 30 Mbps. As highlighted previously, a uniform random

generator is used to generate the bandwidth of the flows between 128 Kbps and 4.096 Mbps. Using the uncorrelated class of weighing, the price of the randomly generated flows are calculated.

After the network topology is deployed, using the multipath routing model and Dijkstra's algorithm mentioned in 3.4.2 subsection, k ($k = 2$ in this scenario) shortest paths from each source-destination is computed. This is shown in Figure 3.4. Using these pre-computed paths, the optimal solution using the exact method is achieved, the results of which is compared to the solution attained through the GCP and SA algorithms.

As mentioned earlier in this chapter a batch processing CAC is proposed. In order to show that batch processing leads to a considerable degree of improvement in terms of revenue maximisation, the optimal and GCP and SA results are compared to the on-line scheme, where flows are admitted individually. The results in Figure 3.5 shows that batch processing results in a higher revenue gain comparing to the on-line scheme where the decision as to admit or reject is performed for each request independently as they arrive at the gateway. It can also be observed that SA algorithm outperforms GCP algorithm and it achieves a better near-optimal solution. The results also demonstrate that SA is capable of efficiently solving the complex CAC problem with difficult constraints and can be further adopted in scenarios with higher complexity i.e. with more MAPs, flows and *sps*; where an optimal solution cannot be achieved using the exact methods.

3.6 Concluding Remarks

In this chapter, a joint admission control and routing in WMNs is formulated as an MKAR problem, which maximises the total revenue from all the carried connections in the network whilst taking into account the bandwidth constraints of access and backhaul links. First, an optimal pricing policy that considers the complexity of the problem and the connection blocking probability is derived. Using this optimal pricing policy, the proposed revenue-based CAC algorithm is investigated. An optimal solution is found through applying exact methods to the CAC algorithm and results are compared to two sub-optimal solutions that are achieved through incorporating the GCP algorithm and the meta-heuristic search algorithm SA to the CAC algorithm.

The analysis indicates that using the uncorrelated class of weighing as the pricing model results in higher revenue gain, lower blocking probability and less computational complexity. Moreover, the proposed joint CAC and route assignment scheme results in higher revenue gain comparing to the on-line scheme. Comparative simulation studies show that SA algorithm achieves a better near-optimal solution comparing to GCP and have the tendency to be adopted in scenarios with higher complexity.

Chapter 4

Admission Control Scheme for Proxy Mobile IPv6 Networks

4.1 Introduction

Rapid growth in number of mobile devices such as smart phones, PDAs, and laptop computers demand for “anytime anywhere” high-speed Internet access. All-IP mobile networks are a combination of Internet and telecommunication networks which provide better interworking between heterogeneous radio access networks and support ubiquitous communication [20]. Next generation of mobile networks are expected to support a variety of mobile devices over IP based mobile networks. As a result the need to have high-speed Internet access and an intelligent mobility management support along with QoS mechanisms is ever-increasing. In order to provide an efficient mobility support various macro and micro mobility management protocols has been proposed by IETF as outlined in Chapter 2.

As stated in the Background Chapter, PMIPv6 have several advantages over MN based micro mobility protocols such as Cellular IP [21], HMIPv6 [23], FMIPv6 [22] [24] and macro mobility solutions such as MIPv6 [18]. Some of the benefits include minimising the handover latencies, reducing the overheads such as signalling over the wireless link and non-complex deployment given that there is no need to upgrade the protocol stack of MNs [26]. The tunnelling based nature of PMIPv6 as depicted in Chapter 2, dictates that all the traffic regardless of their QoS requirements has to flow through the Local Mobility Anchor (LMA) nodes. Several drawbacks are associated with triangular routing from the CN to the LMA and then to the MN; such as routing issues, tunnelling overheads, and creating congestion around the LMA. These drawbacks can immensely affect the network performance, for instance the bottleneck effect caused by the congestion around the LMA reduces the capacity of the access network and the overall throughput of the network. Thus, managing the resources at the LMA becomes highly important. This becomes the central focus of this chapter and in this respect an admission control scheme that utilises a class-based approach and treats traffic according to their QoS requirements is proposed.

Mobile networks were traditionally only designed for one class of traffic (i.e. voice), but in IP based mobile networks an MN can potentially access a wide range of multimedia rich traffic and multiple classes of service are supported by the networks. Providing differentiated service in order to satisfy QoS requirements of multiple classes of service, demands for new protocols and management schemes. Handling a large number of individual flows and providing them with QoS support is an extremely complex task. Aggregating flows into a limited number of classes according to their QoS requirements helps overcoming this hurdle and simplifies the traffic management task. Taking the class based approach to QoS provisioning, a distinction is made between different classes of traffic accord-

ing to their sensitivity with regards to delay.

In this chapter, in order to address the resource management issues at the LMA, a class-based resource allocation scheme is considered. Real-time traffic is given more priority by being able to simultaneously occupy more resource units, whereas non-real time traffic can use less resource units. As mentioned earlier, the tunnelling based nature of PMIPv6 can cause routing issues, reduce the overall capacity of the network and result in the LMA/s becoming the main bottleneck node/s in the network. Allowing all types of traffic to be routed through the LMA uses the valuable available bandwidth at this node, especially in case of routing non-real time IP traffic, such as Peer to Peer (P2P) downloads, through the LMA. Under this perspective, a cut-down version that does not consider the effect of mobility, the QoS aware Dynamic Route Optimisation (DRO) and routing policy proposed in [49] is implemented, where sessions that are not delay sensitive are to bypass the LMA and establish a direct binding with the CN.

The following issues were addressed through investigating two sets of scenarios:

1. The research work presented considers categorising traffic into two classes: Real-time or non-delay tolerant, and Non real-time or delay tolerant. The PMIPv6 network is modelled as a tandem queuing network, where each node operates as an M/M/m/m queue or Erlang Loss system [39] (Chapter 3, Section 3.4.3), with two types of arrival process. The capacity of each node is assumed to be divided into m resource units.
2. The proposed admission control scheme differentiates between each class of traffic by means of allocating different number of resource units; each real-time session will benefit from the simultaneous use of several resource units whilst each non real-time session can occupy one of the m resource units at a time.
3. The state transition diagram of one node is analysed. The total blocking

probability, and per class blocking probabilities are formulated. Through extensive simulation studies it has been shown that the simulation results verify the mathematical analysis.

4. In the first scenario, sessions from both classes of traffic traverse through the LMA node which is in tandem with the MAGs. Simulation results show that through incorporating the proposed class-based CAC scheme, the performance of the system in terms of total blocking probability as well as per class blocking probability is significantly improved.

5. In the second scenario, a cut-down version of the RO and policy proposed in [49] is implemented, where real-time traffic are routed through the LMA and non real-time traffic are routed directly to the MAGs. Thorough investigation of the performance of the joint DRO and proposed admission control scheme was carried out in MATLAB.

The remainder of this chapter is organised as follows. The following section outlines the background on routing in PMIPv6 and related literature to the routing policy implemented in this chapter (i.e. the cut-down version of DRO). Section 4.3 introduces the problem in detail. PMIPv6 is modelled as tandem queuing network in 4.4, and the state transition diagram is illustrated. A detailed description of the investigated simulation scenarios along with the proposed CAC scheme is given in 4.5. Section 4.6 investigates the performance of the proposed scheme; results of the performance evaluation are also presented in this section. Section 4.7 concludes the chapter.

4.2 Background Study

The routing solution implemented in the second scenario of this chapter is based on the approach proposed in [49], where a routing policy that works in conjunction with a QoS aware DRO is proposed. In this respect, first a brief overview of QoS aware DRO is given; the details of the routing policy are then stated.

A thorough description of the PMIPv6 is given in Chapter 2. It is worth mentioning some of the features of the two functional entities in PMIPv6 here again. The LMA has the same role as an MA and manages the MN's BUs. The MAG's main duty is to perform mobility-related signalling as well as detecting and performing handovers on behalf of all the MNs that are attached to its access link [3] [26]. The MAG is usually implemented on the AR of the MN.

4.2.1 A brief overview of QoS aware DRO and its Policy

QoS aware DRO proposed in [49] aims at extending the RO solutions proposed for HMIPv6 [50][51], where MNs with low mobility and number of sessions bypass the MA and instead directly establish a BU with the CN. In PMIPv6, traffic originated from or sent to an MN has to go through a bidirectional tunnel between a MAG and an LMA. For the uplink direction, traffic initiated from the MN is sent to the MAG, the MAG then tunnels the traffic to the LMA and the LMA sends the data to the CN. Similar principle stands for the reverse direction. This results in traffic traversing through a longer path as oppose to the optimised route, especially when the MN and the CN are topologically closer to each other than to the LMA [78]. Paths containing LMA may become overutilised and paths disjoint from the LMA can be left underutilised. Addressing routing inefficiencies

created by employing tunnelling based PMIPv6 were the underlying motivation involved in proposing the RO and policy in [49].

Currently, it is possible to bypass the LMA for localised routing, when the MN and the CN are attached to access links of the same MAG. The same principle does not hold for the case where the CN and the MN are within two MAGs of the same PMIPv6 domain. The IETF has recently been working in the Network-Based Mobility Extensions (netext) Working Group and has focused on further enhancing the localised routing in PMIPv6 networks in order to facilitate direct binding between MAGs of the same domain [2].

In PMIPv6 each MAG has a globally routable Proxy-CoA, which is static and is the same for all the MNs within that MAG. In order to facilitate the direct communication of MNs and CNs, a MAG needs to have a global address for each of the MNs. To resolve this issue and overcome the security issues, in [49] a Network Address Translation [79] like scheme is implemented, where the MAG can communicate with the CN on behalf of the MN through Proxy Local Care of Address (P-LCoA). These addresses are mapped to the MN ID and are stored in a binding cache of each MAG. Through this addressing mechanism, the authors proposed an RO scheme for two scenarios when: I) The CN is located within the same PMIPv6 domain. II) The CN is located outside the PMIPv6 domain. The two scenarios are detailed below.

In the first scenario, when an MN initiates a session the MAG can detect whether the CN is located within the same PMIPv6 domain. If the CN resides within the same PMIPv6 domain, the MAG sends signalling messages to the MAG of the CN in order to request P-LCoA. A direct route is established between the MN and the CN as soon as the P-LCoA is received by the MAG of the MN. This is illustrated in 4.1.

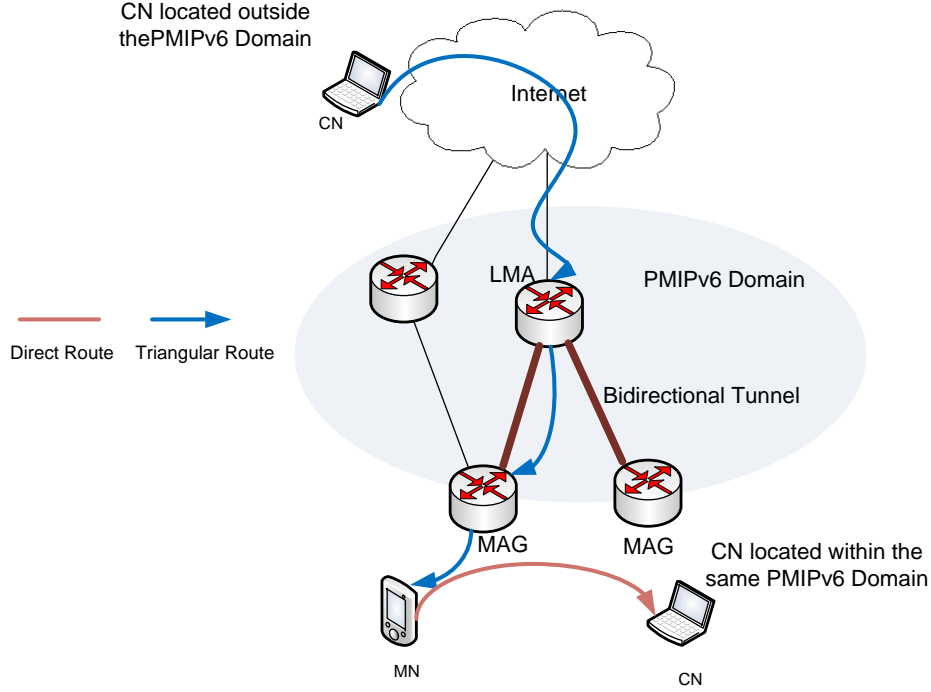


Figure 4.1: Direct and Triangular Routing in PMIPv6

Triangular routing of high bandwidth consuming traffic results in LMAs becoming congested and creating bottleneck effect in the network. This is shown in 4.1, where all traffic destined to the MN are flown from the CN to the LMA, then to the MAG where after decapsulation they are forwarded to the MN. Majority of the high bandwidth consuming traffic are expected to be downloads from servers located outside the PMIPv6 domain, allowing all these background traffic to go through the LMA wastes valuable bandwidth, and introduces more processing alongside tunnelling costs. Therefore in the second scenario, sessions that are not delay tolerant are to be routed through the LMA and session that are delay tolerant are to bypass the LMA given that they are not from a fast moving MN. Mobility rate of the MN was taken into consideration because binding a fast moving MN directly to the CN can result in increased signalling overheads due to

many possible handovers. In this chapter, the effect of mobility is not taken into consideration, and a cut-down version of the QoS aware DRO is implemented as shown in section 4.5.2.

In [49] in order to ensure that low QoS traffic are routed away from the LMA, a routing policy is enforced and implemented on the edge routers of the PMIPv6 networks (i.e. Gateway and MAGs). Looking at their proposed routing policy from the downlink direction, a packet can either be routed to the LMA or it can be routed directly to the MAG. The same holds for the reverse direction. This policy divides the traffic in the network into two: 1) LMA Traffic 2) Non-LMA Traffic. Also, paths that contain an LMA are distinguished from the paths that do not contain the LMA in order to ensure non-LMA traffic avoid the LMA node. As a result there is a set of paths that are prioritised for the LMA traffic, and the other set of paths are not left underutilised.

4.3 Problem Description

With the advances in communication technology, mobile devices have become one of the most important areas in Internet growth. They can access a wide range of multimedia traffic with different QoS requirements. PMIPv6 provides mobility to the IP devices without their involvement, and it can achieve this by relocating the mobility management functions from the MN to the network. Traffic originated from or sent to an MN is routed through tunnels between the MN's corresponding MAG and LMA or vice versa. This results in high bandwidth consumption and bottleneck effect as well as increased processing time at the LMA. Considering this and the huge expected rise in mobile IP traffic in future, the need for employing an efficient resource management scheme in PMIPv6 networks is increasing. Two approaches can be used to address the issue of resource management in PMIPv6 networks: 1) An effective admission control scheme 2) An optimal routing solution. The motivation in this chapter is based on the idea that these two approaches should be jointly considered.

As stated earlier, future IP based mobile networks support a range of multimedia rich traffic with varying QoS requirements. Therefore, if all the traffic originated from the MN are treated the same it may cause QoS disruption and user dissatisfaction. In this chapter the traffic is considered to be either real-time (non-delay tolerant) or non real-time (delay tolerant). The admission control proposed assigns higher priority to the real-time traffic by allocating more resource units to this type of traffic. Capacity of each node is assumed to be divided into m resource units and the admission control scheme works as follows: real-time traffic can benefit from simultaneous use of several resource units and non real-time traffic can occupy fewer resource units.

Routing inefficiencies caused by triangular routing or tunnelling based nature of PMIPv6 can immensely affect the processing time and the bandwidth consumption at the LMA node. It may result in traffic travelling through longer paths as oppose to the optimal path, also certain paths that contain LMA nodes may become highly congested whilst disjoint paths from the LMA may be left under-utilised. As mentioned previously, the LMA itself may become the main bottleneck in the network and reduce the overall throughput of the network. To this end, a cut-down version of the routing solution proposed in [49] is implemented on top of the admission control scheme. Bypassing the traffic that does not require a high level of mobility support away from the LMA, combined with the proposed class-based resource allocation scheme can provide better traffic management as well as differentiated service, and considerably reduce the congestion in the network.

The aim of the proposed solution in this chapter is to address the issue of resource management in PMIPv6 networks which is divided into two parts. First, a class-based admission control is proposed where real-time traffic are given more priority by being assigned more resource units and non real-time traffic are allowed to occupy fewer resource units. In the second part of the solution, a modified version of the routing approach proposed in [49] is implemented on top of the admission control scheme and the performance of the joint routing solution and proposed admission control is investigated.

4.4 System Model

This section sets out a formal definition of the analytical model used in the performance evaluation section. In this chapter, PMIPv6 network is modelled as a tandem queuing network. Capacity of each node is assumed to be divided into m resource units which is equivalent to m servers and each node operates as an independent M/M/m queue or Erlang Loss system [39] (Chapter 3, Section 3.4.3). The primary assumption in future all-IP networks is that n classes of traffic are present, but as mentioned earlier aggregating flows into a few number of classes according to their QoS requirements is much simpler traffic management task than providing QoS for each of the n individual flows. Moreover, with majority of today's Internet traffic being background and being less sensitive to delay (such as downloads of games, music, videos and etc) [80][81], it is wise to have a more general view when it comes to differentiating the classes of traffic in the network. In this respect, two classes of traffic are considered as follows: I) Delay tolerant or non real-time traffic and II) Non-delay tolerant or real-time traffic. A similar distinction between different classes of traffic was made in [49] and [82].

Sessions from both classes of traffic have to traverse through the LMA node and if both classes of traffic are treated the same, it may result in QoS disruption. The proposed admission control scheme in this chapter, takes a class-based approach when it comes to resource allocation and operates as follows: assuming that the bandwidth required by the real-time traffic is much larger than the bandwidth required by the non real-time traffic, each real-time session will benefit from the simultaneous use of several resource units whilst each non real-time session can occupy one of the m resource units at a time.

The amount of bandwidth required by the real-time and non real-time traffic is a and b Kbps respectively. Let's assume that the capacity of all nodes in the network is equal to each other and is denoted by C_T . Assuming that $a \gg b$, then it can be stated that the total bandwidth in each node is equal to m servers or resource units, m being equal to η as shown in equation 4.2. In the proposed model in this chapter, admitting a non real-time traffic requires one of the η resource units and admitting a real-time traffic requires the simultaneous use of several resource units. Number of required resource units by the non real and real-time traffic are shown by m_b and m_a respectively, where $m_b = 1$ and $m_a = a/b$. Considering one class of traffic at a time and assuming that there is no arrival from the other class, it can be stated that the number of real-time traffic that can be admitted at each node is equal to γ and the number of non real-time traffic that can be admitted at each node is equal η such that:

$$\gamma = C_T/a \quad (4.1)$$

$$\eta = C_T/b \quad (4.2)$$

One of the primary assumptions in this chapter is that there is an identical service time offered in each node of the network with higher service time for the real-time traffic, this assumption restores the independence of inter-arrival times and packet lengths; hence the Kleinrock independence approximation is valid and nodes in the network can be modelled as independent M/M/m/m queuing systems [39] (Chapter 3).

Another valid assumption is the offline route discovery; i.e. admitting the sessions at the gateway is performed with the prior knowledge of the bottlenecks in the network. Dijkstra algorithm is used to find the set of shortest paths in the

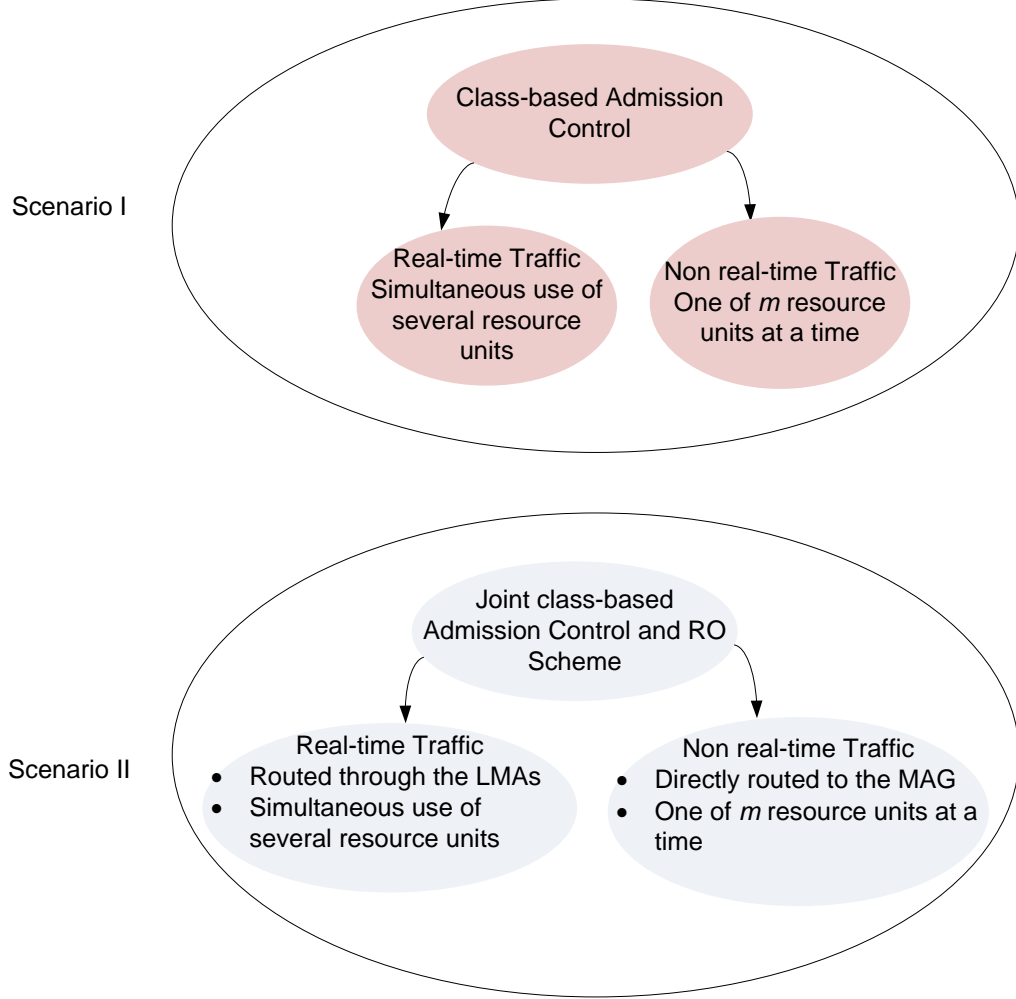


Figure 4.2: Overview of the proposed resource management scheme for PMIPv6

network. In this chapter two scenarios are considered; it is essential to mention that the proposed admission control scheme is performed on the downlink traffic in both scenarios. First, all sessions are to be routed through the LMA node whilst using the proposed admission control scheme. Hence, in the first set of scenario the major bottleneck in the network is the LMA itself. As stated earlier, another way of tackling the issue of resource management in PMIPv6 networks is through addressing the routing inefficiencies caused by the salient features

of these networks. In this respect, implementing a routing solution on top of the proposed admission control is pursued in the latter scenario. A thorough description of the routing solution proposed in [49] is given in section 3.2. In the second set of scenario, a modified version of their solution is implemented which works in conjunction with the proposed admission control scheme. In the modified version the mobility rate of the MN is not taken into consideration and the implemented routing solution in a nutshell operates as follows: upon arrival of a flow at the gateway, a decision as to route this flow through the LMA or the MAG based on the class of the traffic is made. Real-time traffic is routed through the LMA and non real-time is directly routed to the MAG.

An overview of the proposed admission scheme and the two different scenarios are shown in Figure 4.2. A detailed explanation of the above mentioned analytical model is outlined in the next subsections.

4.4.1 Network Model

The network model used for both sets of scenarios is outlined in this section. A PMIPv6 network is considered as a connectivity graph $G(V; E)$, where V is a set of routers and E is the set of links in the network. K is the set of LMA routers in the PIMIPv6 network, M is the set of MAGs (or ARs) and g is the gateway router such that $M \in V - \{K, g\}$. Set of all of the shortest paths from the gateway to each MAG is defined as P . It is essential to distinguish between the paths that contain an LMA and the paths that do not contain an LMA. In this respect, let $P_{g,k}^{LMA}$ and $P_{k,m}^{MAG}$ be the set of all paths from gateway g to the LMA $k \in K$ and from LMA $k \in K$ to MAGs $m \in M$ respectively, which define the set of paths that contain an LMA and is notated as \tilde{P} . The set of \bar{P} paths defined as

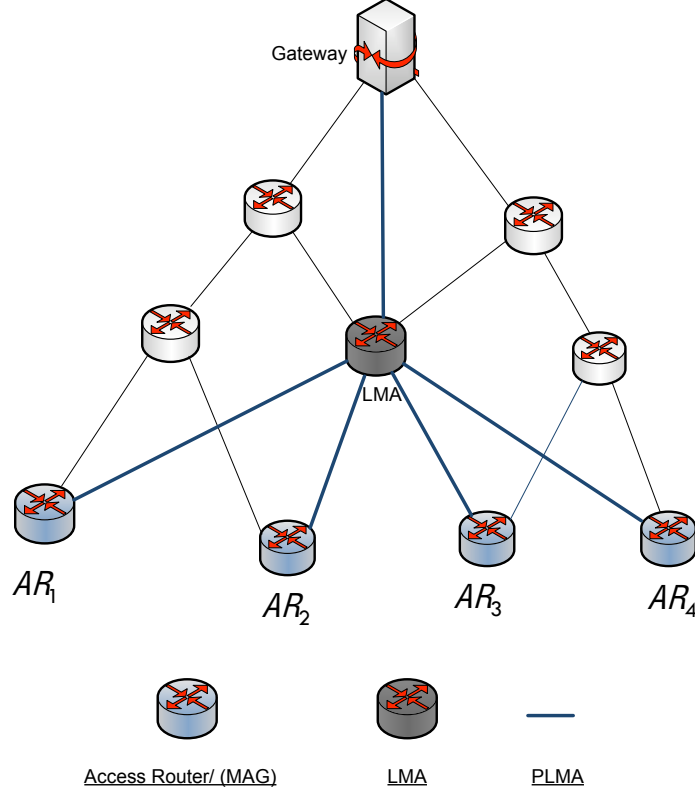


Figure 4.3: Topology for Scenario I

$\bar{P} \in P - \tilde{P}$, is the set that does not contain an LMA. The session arrivals for both classes of traffic ($n = 2$) is assumed to follow the Poisson distribution with an average arrival rate λ_n and exponentially distributed service time with mean $1/\mu_n$. Service time offered in all of the nodes in the network is identical and the service time offered to the real-time traffic is higher; i.e. if $\mu_1 = \mu$ then $\mu_2 = 5\mu$.

The topologies used in the two scenarios are shown in Figures 4.3 and 4.4.

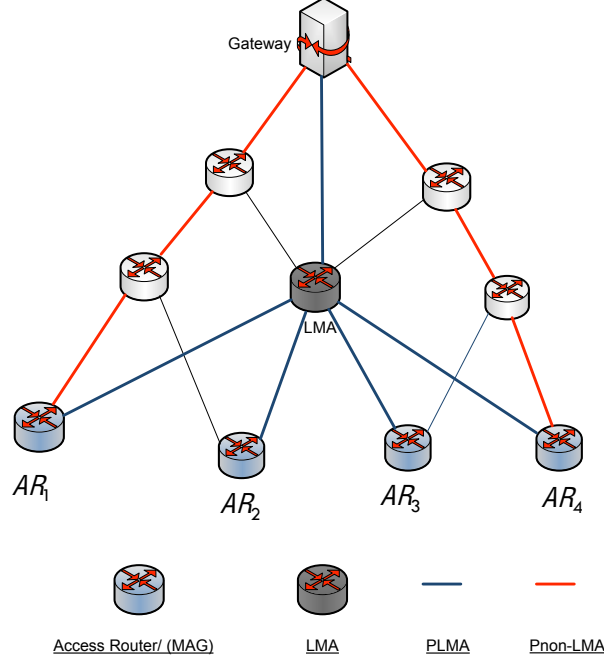
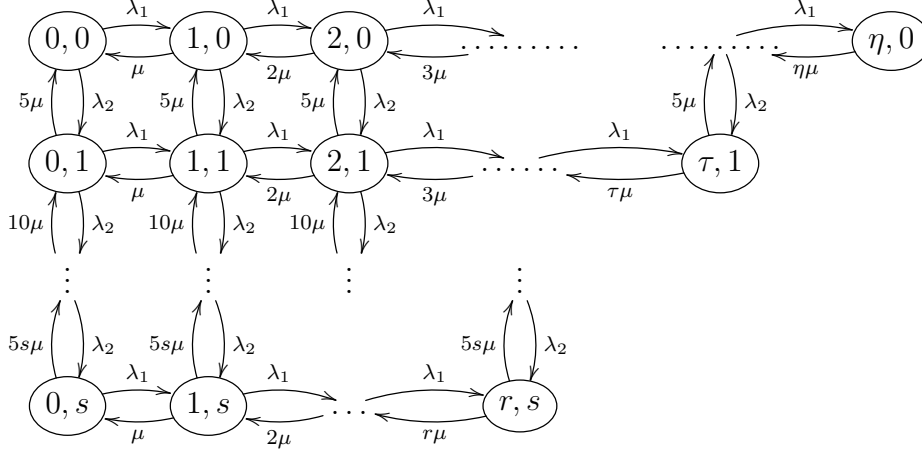


Figure 4.4: Topology for Scenario II

4.4.2 State Transition Diagram

In this section the state transition diagram for one node in the first scenario is analysed. State (i, j) means that i requests of non real-time and j requests of real-time class are present in the system. Let λ_1, λ_2 and $\mu, 5\mu$ be the average call arrival and mean service time of non real-time and real-time traffic respectively. Consider an LMA node within a PMIPv6 network with capacity C_T , this can be analysed by a 2-dimensional Markov chain [83] (Chapter 3, Section 3.4.3), as follows:

4.4. System Model



Let $P(i, j)$ be the steady-state probability of system being in state (i, j) . The probability of system being in an equilibrium state can be found by solving the general balance equation (flow in = flow out) for an internal state:

$$\begin{aligned} \lambda_1 p(i-1, j) + \lambda_2 p(i, j-1) + (i+1)\mu p(i+1, j) + (j+1)5\mu p(i, j+1) \\ = j5\mu p(i, j) + i\mu p(i, j) + \lambda_2 p(i, j) + \lambda_1 p(i, j) \end{aligned} \quad (4.3)$$

and by the method suggested in [83] (Chapter 3), this is shown below :

$$P(i, j) = \frac{\frac{1}{i!} \left(\frac{\lambda_1}{\mu}\right)^i \frac{1}{j!} \left(\frac{\lambda_2}{5\mu}\right)^j}{\sum_{j=0}^{\eta} \frac{1}{j!} \left(\frac{\lambda_2}{5\mu}\right)^j \sum_{i=0}^{\eta-jm_a} \frac{1}{i!} \left(\frac{\lambda_1}{\mu}\right)^i} \quad (4.4)$$

In the state diagram, horizontal arrows to the right and left correspond to arrival and departure of non real-time traffic into the system. It can be concluded that a non real-time traffic is blocked if all the η servers are busy and this occurs when the system is at the rightmost of any row. Using equation 4.4 and summing the probabilities of the $s+1$ states gives the blocking probability of this class of traffic, where $P_{b_{non-rt}}$ denotes the probability of a non real-time arrival being

blocked or denied access to the LMA:

$$\begin{aligned}
 P_{b_{non-rt}} &= p(0,0) \left[\frac{1}{\eta!} \left(\frac{\lambda_1}{\mu} \right)^\eta + \frac{1}{(\eta - m_a)!} \left(\frac{\lambda_1}{\mu} \right)^{\eta - m_a} \left(\frac{\lambda_2}{5\mu} \right) \right. \\
 &\quad \left. + \dots + \frac{1}{(\eta - sm_a)!} \left(\frac{\lambda_1}{\mu} \right)^{\eta - sm_a} \frac{1}{s!} \left(\frac{\lambda_2}{5\mu} \right)^s \right] \\
 P_{b_{non-rt}} &= \frac{\sum_{j=0}^s \frac{1}{j!} \left(\frac{\lambda_2}{5\mu} \right)^j \frac{1}{(\eta - jm_a)!} \left(\frac{\lambda_1}{\mu} \right)^{\eta - jm_a}}{\sum_{j=0}^s \frac{1}{j!} \left(\frac{\lambda_2}{5\mu} \right)^j \sum_{i=0}^{\eta - jm_a} \frac{1}{i!} \left(\frac{\lambda_1}{\mu} \right)^i} \quad (4.5)
 \end{aligned}$$

Vertical arrows up and down, represent arrival and departure of real-time traffic into the node. A real-time arrival is blocked if at least $\tau + 1$ servers are busy, where $\tau = \eta - m_a$, in other words when less than m_a resource units are idle or available. In order to calculate the blocking probability of real-time traffic $P_{b_{rt}}$, $\eta + 1$ probabilities are summed as follows:

$$\begin{aligned}
 P_{b_{rt}} &= \sum_{i=\eta - m_a + 1}^{\eta} p(i, 0) + \sum_{i=\eta - 2m_a + 1}^{\eta - m_a} p(i, 1) + \\
 &\quad \dots + \sum_{i=\eta - sm_a + 1}^{\eta - (s-1)m_a} p(i, s-1) + \sum_{i=0}^r p(i, s)
 \end{aligned}$$

The probability that traffic of type real-time is blocked from the system is:

$$\begin{aligned}
 P_{b_{rt}} &= \frac{\sum_{j=0}^{s-1} \frac{1}{j!} \left(\frac{\lambda_2}{5\mu} \right)^j \sum_{i=\eta - (j+1)m_a + 1}^{\eta - jm_a} \frac{1}{i!} \left(\frac{\lambda_1}{\mu} \right)^i}{\sum_{j=0}^s \frac{1}{j!} \left(\frac{\lambda_2}{5\mu} \right)^j \sum_{i=0}^{\eta - jm_a} \frac{1}{i!} \left(\frac{\lambda_1}{\mu} \right)^i} + \\
 &\quad \frac{\frac{1}{s!} \left(\frac{\lambda_2}{5\mu} \right)^s \sum_{i=0}^r \frac{1}{i!} \left(\frac{\lambda_1}{\mu} \right)^i}{\sum_{j=0}^s \frac{1}{j!} \left(\frac{\lambda_2}{5\mu} \right)^j \sum_{i=0}^{\eta - jm_a} \frac{1}{i!} \left(\frac{\lambda_1}{\mu} \right)^i} \quad (4.6)
 \end{aligned}$$

4.5 Simulation Scenarios

This section provides a detailed explanation of the two studied simulation scenarios and the proposed admission control scheme.

4.5.1 First Scenario: All through the LMA

In this scenario, sessions from both classes of traffic have to traverse through the LMA node. This means that the set of paths that contain the LMA node are used and set of \bar{P} paths are left underutilised. Nodes within the network are modelled as z independent two-dimensional M/M/m/m queues, this is illustrated in figure 4.5. Each node can be analysed by the same state transition diagram illustrated in the previous section.

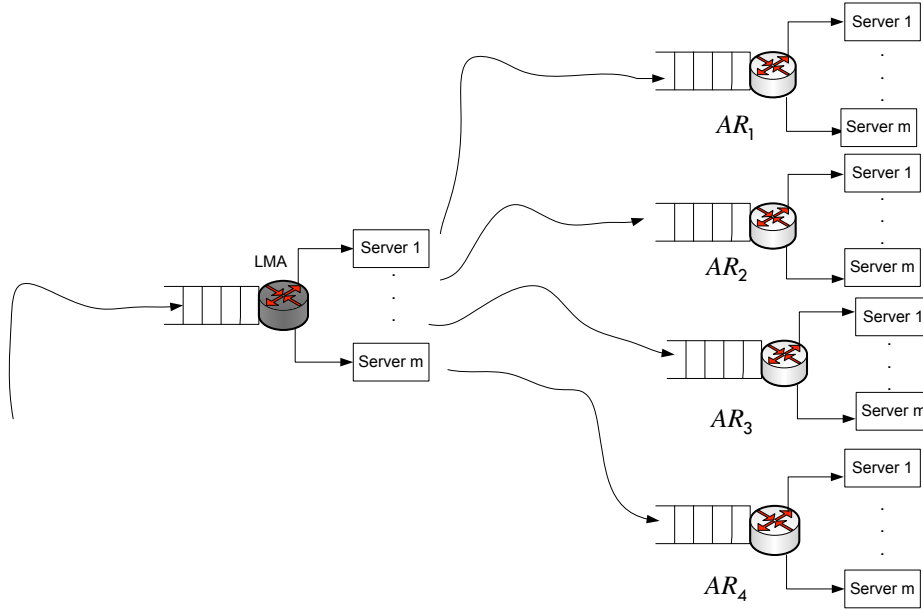


Figure 4.5: Scenario I: z independent M/M/m/m queues

Looking at the system from the downlink and considering Figures 4.3 and 4.5, it can be concluded that the only bottleneck in the system is the LMA node and each session has to pass through two independent queues in tandem before it reaches its destination. This means that a session travels one of these four routes: LMA-AR₁, LMA-AR₂, LMA-AR₃, or LMA-AR₄. The admission control scheme for the downlink works as explained below, the same principle stands for the reverse direction (uplink):

- Upon arrival of a new session x at the gateway, the class and destination of the session are distinguished. Destination of session x can be one of the four ARs, i.e. $D_x \in \{AR_1, AR_2, AR_3, AR_4\}$ and class of session x can be either non real-time or real-time.
- The decision as to admit or block the session at the LMA is made on the basis of the number of available resource units and number of required resource units.
- After the session is admitted at the LMA, it occupies one or m_a of the η resource units depending on its class during the session's life-time (till the end of its service time).
- If enough resource units are available in the second or last node, the session will be forwarded to the second node and remains in one or m_a of the η servers of the second/last node till its service time is finished. Otherwise, the session gets blocked.

Figure 4.6 describes the decision process of the proposed class-based admission control.

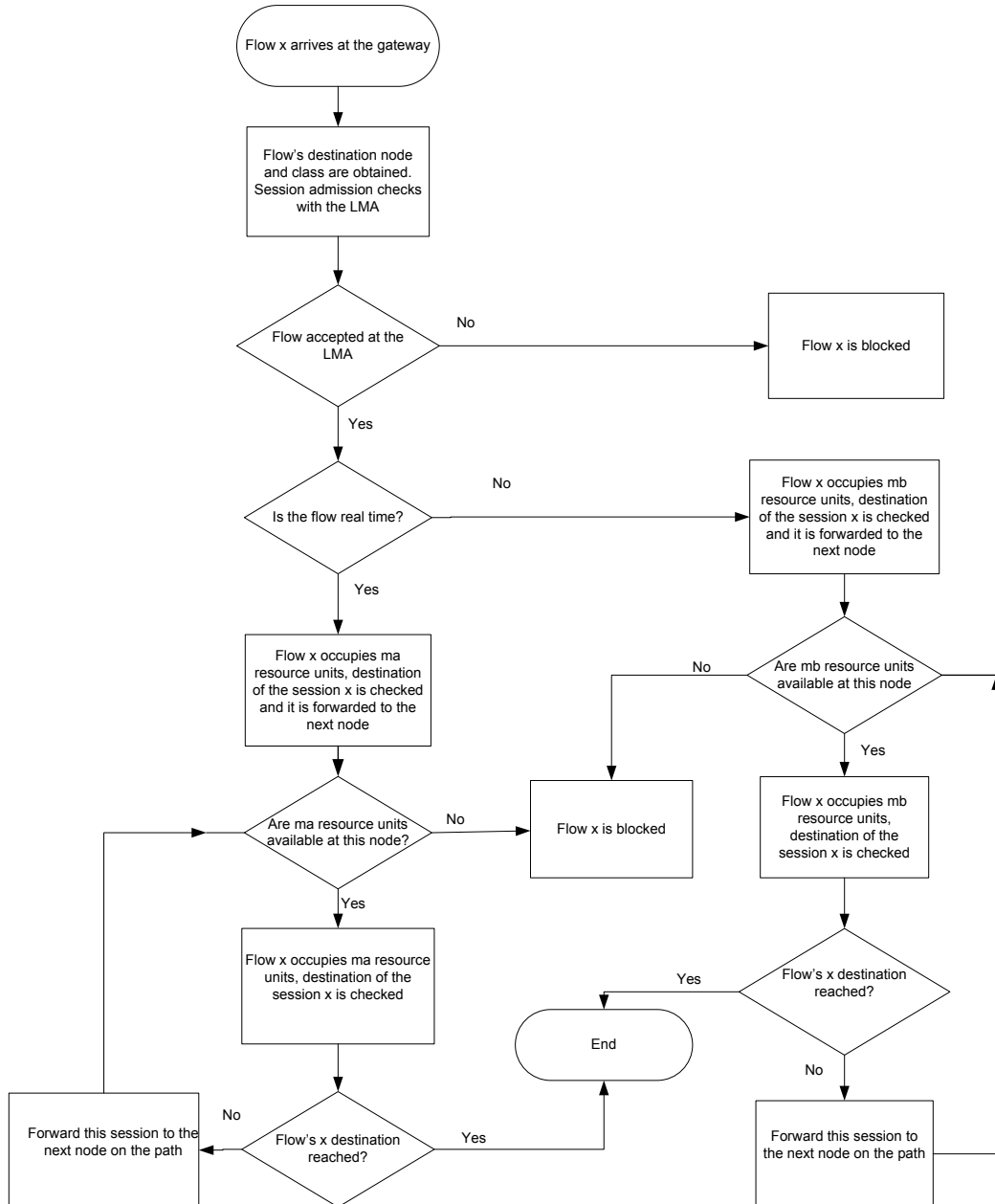


Figure 4.6: Class-based Admission Control Decision Making Process

4.5.2 Second Scenario: Using RO schemes

An RO scheme similar to the QoS aware DRO explained in section 4.2, is implemented that operates alongside the proposed admission control scheme from the first scenario. Session's mobility rate is not taken into consideration, and the routing policy works as follows: Upon arrival of a flow at the gateway, based on the flow's class of the traffic, a decision as to route the flow directly to the specific MAG or the LMA/s is made. If the session is non real-time and it is decided to be routed directly to the MAG, the least cost path from the \bar{P} set of paths is calculated and the non real-time flow is routed through the least cost path. Calculating the least cost path for a real-time session is done in two stages: 1) The least cost path from the gateway to the LMA k is calculated and the flow is routed along this path till it reaches LMA k . 2) The least cost path from the LMA k to the specific MAG is calculated and the flow is routed through this path.

Looking at the joint admission control and RO mechanism from the downlink direction, it can be stated that though sessions from each class of traffic are routed through two separate paths, they still share resource units at the MAGs. Therefore, entry nodes to the network as well as nodes along the path toward the MAGs, can be modelled by independent one-dimensional M/M/m/m queues and MAGs can be modelled as independent two-dimensional M/M/m/m queues as illustrated in section 4.4.2. It must be highlighted that at all the nodes in the network the value of m_b and m_a is the same as it was in scenario one. Depending on the class and the destination of a flow the first entry point to the network can either be one of the LMAs or the first node of one of the \bar{P} paths. Once the session admission has checked that enough resource units are available, the session is admitted at one of these two nodes, and then it is processed and forwarded to

the next node along the path until it reaches the MN at one of the serving ARs (or MAGs).

Real-time traffic has to go through one of the LMA path (i.e. \tilde{P}) to the destination (or to the specific MAG). At each node along this path the number of available resource units are checked and if less than m_a resource units found to be idle the session is blocked. The same holds for the non real-time traffic traversing through non-LMA (i.e. \bar{P}) paths. Once flows from different classes of traffic reach a MAG that their MNs are attached to, then the problem of limited resource units and efficiently rationing them amongst two classes of traffic becomes apparent again. The proposed admission control from scenario one ensures resource units at the MAGs are rationed depending on the QoS requirements of the traffic, i.e. assigning m_a resource units to the real-time and m_b resource units to the non real-time traffic.

4.6 Performance Evaluation

The proposed solutions in this chapter are further investigated by means of MATLAB simulations. The main focus is on the results of simulations carried out in the MATLAB but first results of the mathematical analysis are presented to verify the simulation environment built for one node.

4.6.1 Blocking Probability of One Node

In this section only one node is considered and the results of the analytical model mentioned in section 4.4.2 are compared to the simulation results. The capacity

of the LMA node is assumed to be $C_T = 5$ Mbps. A uniform random generator is used to generate the flow or session arrivals, bandwidth of each real-time and non real-time traffic is considered to be as follows: $a = 500$ Kbps and $b = 100$ Kbps respectively. Here η is equal to 50, $m_a = 5$ and $m_b = 1$. This means that each real-time traffic will access the five resource units simultaneously whereas each non real-time traffic only occupies one resource unit at a time. Statistics show that large part of traffic in the Internet is associated to the P2P downloads [81], while a previous report by Cisco mentioned 70% of traffic is associated to the P2P downloads [80]. Therefore, in these simulation scenario 70% of the total arrivals are non real-time traffic. Simulation results illustrated in Figure 4.7 are achieved by running 100 rounds of simulations using different seeds.

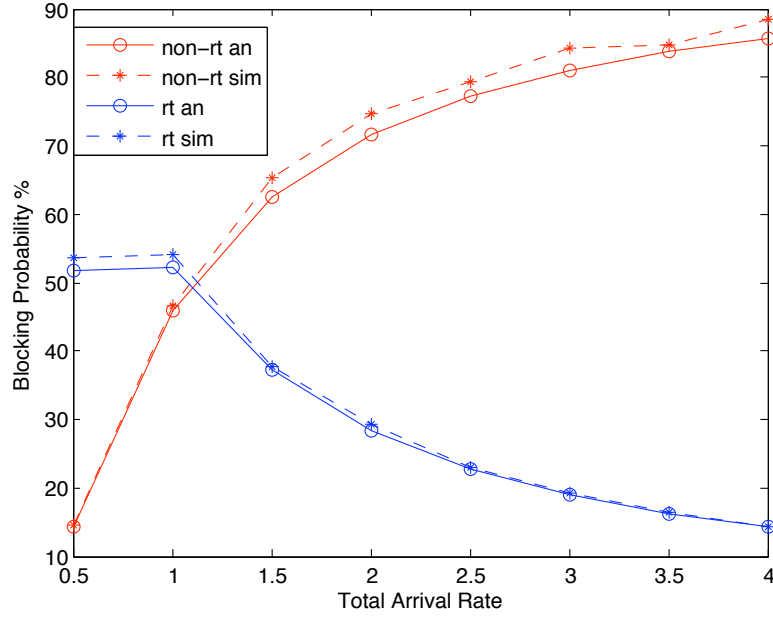


Figure 4.7: Blocking Probability vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points

Figure 4.7 shows the simulation and analytical results on the same plot, while

their difference is less than 5 percent. This verification provides confidence to further extend the simulation environment to a tandem queuing network. The high blocking probability of real-time traffic in the arrival rate range of 0.5 and 1, is due to only 30% of this low total arrival rate being real-time.

4.6.2 Blocking Probability Scenario I

To investigate the impact of the proposed admission control scheme on the total and per class blocking probability, the topology in Figure 4.3 is considered. There are one gateway node, four AR nodes (MAGs) and five intermediate nodes including the LMA node that provide the backhaul routing in the topology. The capacity of a node is denoted by C_T and each node is assumed to have equal capacity $C_T = 20$ Mbps. Similar to previous section a uniform random generator is used to generate the session or flow arrivals, bandwidth of each real-time and non real-time traffic is considered to be as follows: $a = 500$ Kbps and $b = 100$ Kbps, respectively. Here η is equal to 200, $m_a = 5$ and $m_b = 1$. Once the network topology is deployed based on the Dijkstra's algorithm [39] (Chapter 5, Section 5.2.3) and incorporating the approach proposed in [74], P shortest paths from the gateway to each of the four ARs are computed. Set of \tilde{P} paths that contain the LMA node are distinguished from the P pre-computed shortest pathes. Looking at the system from the downlink and referring to Figure 4.3 and 4.5, it can be stated that each flow arrival at the gateway has to go through 2 nodes till it reaches the MN at the one of the serving ARs (or MAGs) and that the LMA node is in tandem with the other four ARs.

The total arrival rate in the simulation scenarios is $\lambda_1 + \lambda_2$. Similar to the previous section, each result is produced after running the simulation 100 times using

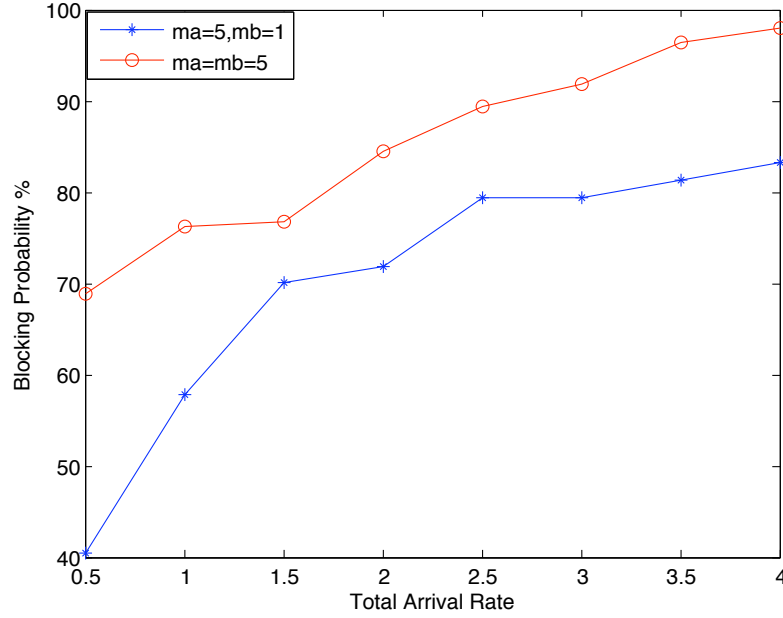


Figure 4.8: Total Blocking Probability vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points

different seeds. First, the total blocking probability for each arrival rate is computed. The results are then compared to the method where there is no distinction between real-time and non real-time traffic in terms of number of resource units used i.e. $m_b = m_a = 5$, and sessions from both classes of traffic require the simultaneous use of five resource units at a time. As displayed in Figure 4.8, it can be gathered that the proposed admission control scheme results in a significant decrease in total blocking probability comparing to the case where sessions from both classes of traffic are treated the same. Moreover, as the session arrival rate increases and as a result the congestion level in the network becomes higher, the gap between the blocking probability of the proposed admission control and the case where $m_b = m_a = 5$ becomes larger and the proposed admission control scheme outperforms the conventional method considerably.

Next, blocking probability per class of traffic using the proposed admission control is attained. The results are then compared with the scenario where $m_b = m_a = 5$. Results in Figure 4.9 demonstrate that by using the proposed admission control, blocking probability of non real-time traffic has lowered. This was expected as in the proposed scheme each non real-time traffic requires one resource unit at a time at each node, whereas with no distinction between real and non real-time traffic each non-real flow will simultaneously occupy 5 resource units at each node though it does not require that many resource units.

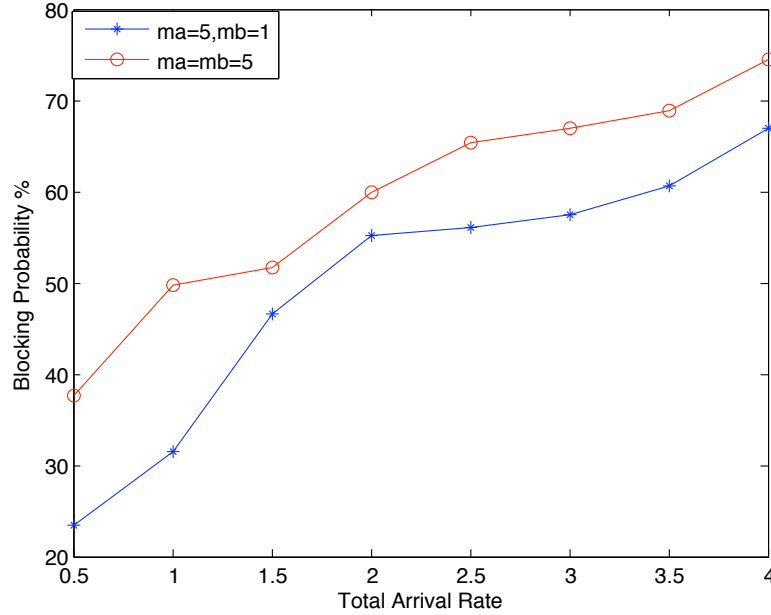


Figure 4.9: Blocking Probability of non real-time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points

The blocking probability of real-time class of traffic is computed, results of which are compared to the case where $m_b = m_a = 5$ in Figure 4.10. The proposed admission control scheme results in reduction of blocking probability of real-time traffic, especially under high congestion in the network. As the total session

arrival rate increases, the number of arrivals from the non real-time class of traffic increases 70% more than the real-time class of traffic. As a result, $P_{b_{nrt}}$ increases and $P_{b_{rt}}$ decreases as the total arrival rate increases. Overall it can be concluded that by treating the two classes of traffic differently a great reduction in both total and per class blocking probabilities is achieved.

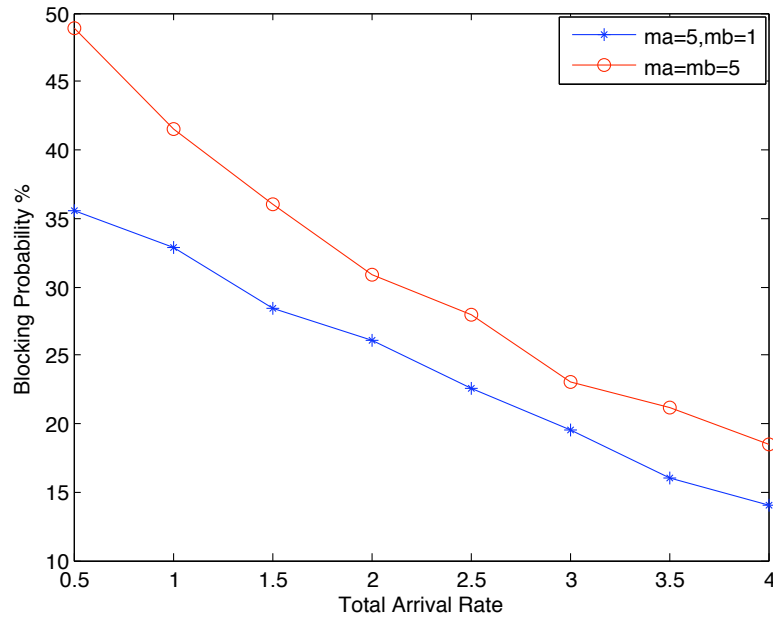


Figure 4.10: Blocking Probability of real-time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points

4.6.3 Blocking Probability Scenario II

The results in the previous section demonstrate the impact that the proposed admission control has on total blocking probability as well as per class blocking probability. In this section the joint performance of the proposed admission control and the modified version of QoS aware DRO as outlined in section 4.2

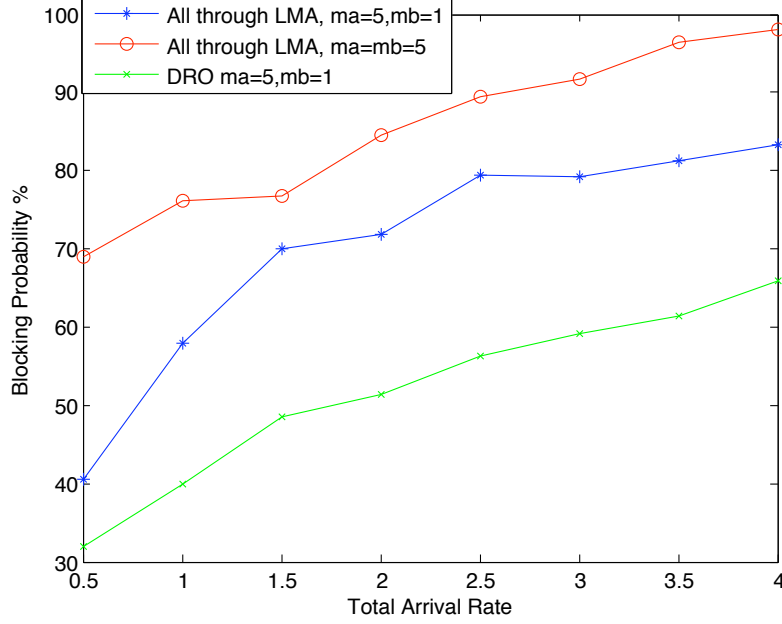


Figure 4.11: Total Blocking Probability vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points

is investigated. The objective of employing the QoS aware DRO in conjunction with the proposed admission control scheme is to lower the blocking probability by distributing the traffic through the underutilised paths in the network. The topology in Figure 4.4 is considered and similar to scenario one, each node is assumed to be equal in terms of capacity and $C_T = 20$ Mbps. Flow arrival generation, and the bandwidth required by each class of flow is the same as the previous scenario. Also, η is equal to 200, and $m_a = 5$ and $m_b = 1$. Path construction also follows the same principal as scenario one, but in this scenario set of paths containing an LMA node and containing no LMA node are distinguished from each other and both (\tilde{P}, \bar{P}) are utilised. Refereing to Figure 4.4, each real-time session has to go through one of the \tilde{P} paths shown in blue. Each non real-time session goes through one of the \bar{P} , shown in red. It is worth mentioning that

similar to the two previous sections, all the results illustrated in this section are obtained through running the simulation 100 times using different seeds.

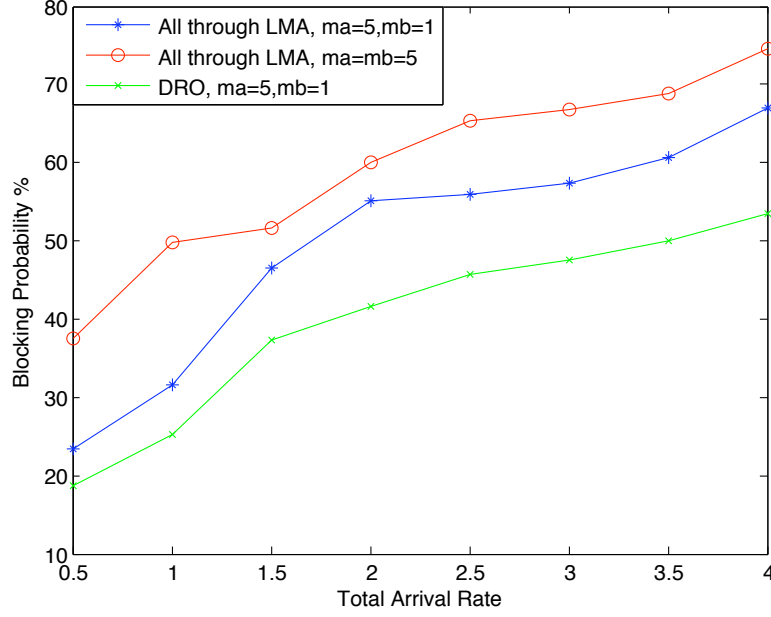


Figure 4.12: Blocking Probability of non-real time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points

Using DRO, the non real-time traffic which forms the majority of the traffic is routed away from the LMA and routed through the underutilised paths. The total blocking probability is computed, results of which are compared to two sets of results previously obtained: the results of scenario one and to the case where all the traffic is routed through the LMA and no distinctions between real-time and non real-time traffic in terms of resource units allocation is considered (i.e. $m_b = m_a = 5$). As shown in Figure 4.11, it can be stated that incorporating DRO in conjunction with the proposed admission control results in a significant degree of improvement in terms of reducing the total blocking probability. Furthermore, from Figure 4.11 it can be gathered that as the arrival rate increase and the congestion level in the network becomes higher, the curve of total blocking

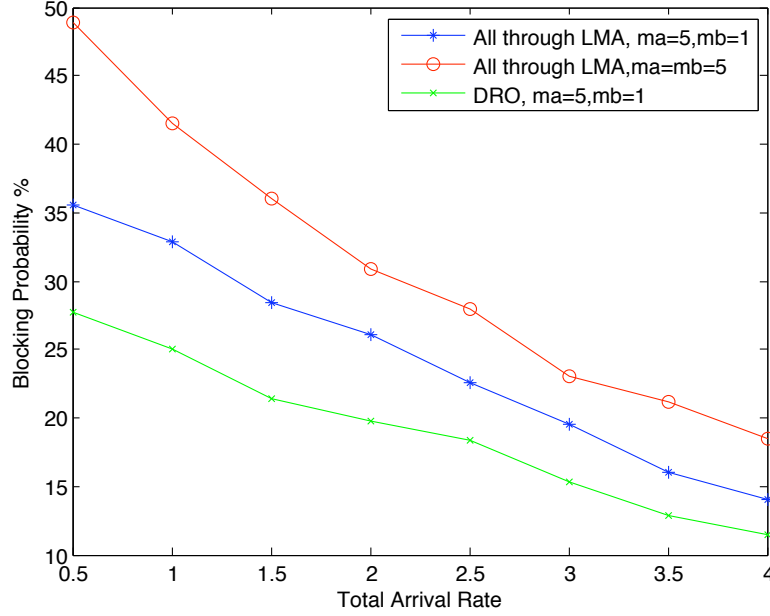


Figure 4.13: Blocking Probability of real-time traffic vs arrival rate, where arrival rate is total arrival rate of real-time and non real-time traffic, with 70 percent of the total arrival rate being non real-time at all the points

probability achieved through using the DRO jointly with the proposed admission control moves further away from the other two curves because the LMA is no longer the main bottleneck node as the traffic is distributed throughout the network.

Blocking probability per class of traffic is also attained through simulations. The results are compared to the results achieved in scenario one and to the case where all the traffic is routed through the LMA and no distinctions between real-time and non real-time traffic in terms of resource units allocation is considered (i.e. $m_b = m_a = 5$). Figure 4.12 illustrates that by using the proposed joint admission control scheme and DRO, blocking probability of non real-time traffic has been reduced. Moreover, under high congestion the joint approach performs much better than the other two, as the blocking probability of non real-time traffic

does not exceed 55 percent.

Incorporating DRO means that only the real-time traffic has the privilege of using the LMA node and this results in lowering the congestion level at the LMA and reducing the probability of the LMA becoming the main bottleneck node in the network. Figure 4.13 illustrates the blocking probability of real-time traffic using three different approaches on the same plot. Real-time blocking probability is at its lowest when using the joint approach. It can be stated that using the joint approach distributes the traffic throughout the network and increases the overall throughput of the network, hence lower total and per class blocking probability is achieved through this approach.

4.7 Concluding Remarks

In this chapter, the issue of resource management in PMIPv6 is addressed with the intention of improving the network performance. First, a class-based admission control scheme is presented to minimise the bottleneck effect caused by triangular routing in PMIPv6. Resource units are rationed amongst different classes of traffic according to their QoS requirements and an analytical model is presented. The PMIPv6 network is modelled as an M/M/m/m tandem queuing network with two types (classes) of arrival process: real time and non-real time traffic. The proposed admission control scheme priorities the real time traffic by allowing this type of traffic to use several resource units at a time. Secondly, in order to address the routing inefficiencies in PMIPv6, a QoS aware DRO is implemented which works in conjunction with the proposed admission control. Non real-time traffic that do not require mobility support mechanisms are routed away from the LMA and are routed through the underutilised paths.

4.7. Concluding Remarks

Mathematical analysis verified the results of extensive simulations of the model built for one node in MATLAB. The model was further extended to a network of queues. Comparative simulation studies demonstrate that incorporating the proposed admission control scheme results in a lower total blocking probability as well as per class blocking probability. Moreover, using the QoS aware DRO in conjunction with the proposed admission control results in a great reduction of total blocking probability, as well as a noticeable reduction of per class blocking probability.

Chapter 5

Conclusions and Future Research

5.1 Concluding Remarks

The main subject studied in this thesis is designing Call Admission Control (CAC) mechanisms for future IP based multi-hop wireless networks. In this respect, two different classes of approach were adapted which are outlined in Chapters 3 and 4. One with the aim of addressing resource management in multi-hop based networks and the other concentrated on improving the overall network capacity in Mobility Agent (MA) based micro mobility solutions.

Chapter 3 focused on designing a CAC mechanism for WMNs. Admission control and routing are strongly interconnected in WMNs, as the availability of resources depends on joint association of each mesh client with a MAP and multi-hop back-haul routing to the Gateway. The approach used was based on decoupling the process of route setting up between the Gateway and MAPs from the admission control decision process. Once k shortest paths were set up between each MAP-Gateway, route selection/assignment and CAC were jointly considered. Hence,

the problem of joint admission control and route assignment in WMNs was formulated as an MKAR problem. The objective of the optimisation problem was to maximise the total revenue from all the carried connections in the network whilst taking the bandwidth constraints of access and backhaul links into account. Impact of implementing different pricing strategies on the total revenue gain and connection blocking probability was investigated. An optimal pricing policy that satisfies all the conditions and does not impose extra complexity on the problem was selected and adopted. Real time solutions of the proposed problem were then achieved through incorporating a heuristic and a meta-heuristic approach. Thorough simulations were carried out in MATLAB platform. It was shown that the performance of the system was noticeably enhanced through batch processing. It was also demonstrated that incorporating meta-heuristic approaches results in a better near-optimal solution.

The issue of designing a CAC mechanism for a specific class of MA based micro mobility solutions referred to as PMIPv6 was investigated in Chapter 4. Presence of MAs or LMAs create major obstacles in designing CAC schemes in PMIPv6 networks. Bidirectional tunnelling between MNs and LMAs wastes the valuable wireless resources and leaves certain paths in the network underutilised. Thus, a class-based admission control was proposed to avoid overutilising resources at the LMA, where resource units were rationed amongst different classes of traffic according to their QoS requirements. A distinction between different classes of traffic was made and two classes of real-time and non real-time traffic were considered. The nodes within the PMIPv6 network were modelled as independent two-dimensional M/M/m/m queues. The aim of the proposed class-based CAC scheme was to offer a differentiated service for real-time and non real-time traffic, therefore more resource units were allocated to real-time traffic at each node. The performance of the proposed class-based CAC schemes has been investigated by

means of simulation. It has been demonstrated that by using the proposed scheme promising benefits in terms of reducing the total blocking probability as well as blocking probability per class of traffic are achieved.

Finally, to allow maximum usage of the network resources in PMIPv6 networks a cut-down version of an RO scheme is implemented. Non real-time traffic that is delay tolerant is routed through underutilised regions of the networks by directly establishing a binding with the CN. Therefore, the bottleneck effect at the LMA is minimised but still two classes of traffic share resources at MAGs. Thus, the RO scheme must operate alongside the proposed CAC in order to route optimise the delay tolerant traffic away from the LMA and efficiently manage resource unit allocation at the MAGs. In this respect, paths that contain LMAs are distinguished from those that do not pass through LMA nodes. Simulation results demonstrated that by implementing the joint routing and admission control scheme, the total blocking probability as well as blocking probability per class of traffic were substantially reduced in comparison to the scenario where only the proposed CAC scheme was implemented and to the scenario where all the traffic was routed through the LMA without applying the class-based CAC scheme.

5.2 Future Research

A substantial amount of open issues and potential avenues for future research exist in this subject. A summary of the issues that relates to the contributions made in this thesis is presented below:

- The joint CAC and routing algorithm proposed in Chapter 3, does not count for interference. Therefore, an interesting dimension is to look into

the inter-flow and intra-flow interference in the route setting up process i.e. when constructing the shortest path for each source-destination. This requires considering both the diversity of channel assignment and the link capacity.

- The proposed CAC scheme for PMIPv6 in Chapter 4, considers a static differentiated resource allocation for the two classes of traffic, where the value for m_a and m_b is fixed. One avenue for future research would be to adopt a dynamic resource allocation scheme where the value of m_a is dynamically changed according to the traffic intensity of the real-time traffic.
- The CAC scheme in Chapter 4 is class-based, but it does not consider bandwidth reservation techniques or in specific resource unit reservation techniques. Per flow bandwidth or resource unit reservation along all the nodes in a path will make the CAC scheme unscalable, as many signalling exchanges need to be made in order to reserve resources. Therefore, implementing a dynamic or fixed resource unit reservation scheme at the bottleneck nodes such as the MAGs or at the LMAs, can be an attractive extension to the scheme proposed in Chapter 4. Different factors such as avoiding inefficient utilisation of system resources should be considered in future exploration along these line.
- The CAC schemes proposed in this thesis did not consider the effect of mobility. A cut-down version of the RO scheme proposed in [49] was implemented in Chapter 4, to ensure that non real-time traffic bypass the LMA. In the modified version as long as an arrival process is classified as non real-time, it is routed through the paths that does not contain the LMA

without taking the mobility rate of the MN into consideration. Hence, a possible area of future research would be to study the effect of mobility at the node level in PMIPv6 networks and either implement or design an RO inspired by the one proposed in [49].

References

- [1] I. Akyildiz and X. Wang, *Wireless Mesh Networks*. Wiley; 1 edition (March 10, 2009), 2009. 13, 27, 29, 30, 31
- [2] I. Soto, C. J. Bernardos, M. Caldern, and T. Melia, “Pmipv6: A network-based localized mobility management solution,” *The Internet Protocol Journal*, vol. 13, Sep 2010. 13, 32, 33, 34, 36, 38, 85
- [3] S. Gundavelli, K. Leung, V. Devarapalli, K. Chowdhur, and B. Patil, “Proxy mobile ipv6,” *IETF RFC 5213*, August 2008. 13, 36, 37, 38, 39, 84
- [4] H. Perros and K. Elsayed, “Call admission control schemes: a review,” *Communications Magazine, IEEE*, vol. 34, pp. 82 –91, nov 1996. 16, 22
- [5] M. Ahmed, “Call admission control in wireless networks: a comprehensive survey,” *Communications Surveys Tutorials, IEEE*, vol. 7, pp. 49 –68, qtr. 2005. 22, 23, 25, 39, 56
- [6] M. Lohi and K. Madani, “Using artificial neural networks to improve performance in a wireless atm network,” in *The 49th IEEE International Conference on Vehicular technology Conference, Houston, Texas,, May 1999*. 22

REFERENCES

- [7] M. Ghaderi and R. Boutaba, “Call admission control in mobile cellular networks: a comprehensive survey,” *Wireless Communications and Mobile Computing*, vol. 6, no. 1, pp. 69–93, 2006. 22, 52
- [8] D. Niyato and E. Hassain, “Call admission control for qos provisioning in 4g wireless networks: issues and approaches,” *IEEENetworks*, vol. 19, pp. 5–11, Septeber/October 2005. 23
- [9] Z. Liu and M. El Zarki, “Sir-based call admission control for ds-cdma cellular systems,” *IEEE Journal on Selected Areas in Communications*, vol. 12, pp. 638 –644, may 1994. 25, 26
- [10] J. Knutsson, P. Butovitsch, M. Persson, and R. Yates, “Downlink admission control strategies for cdma systems in a manhattan environment,” in *Vehicular Technology Conference, 1998. VTC 98. 48th IEEE*, vol. 2, pp. 1453 –1457 vol.2, may 1998. 25
- [11] W. S. Jeon and D. G. Jeong, “Call admission control for mobile multimedia communications with traffic asymmetry between uplink and downlink,” *Vehicular Technology, IEEE Transactions on*, vol. 50, no. 1, pp. 59–66, 2001. 25
- [12] J. Li, S. Wu, and S. Liu, “A new cac scheme based on adaptive sinr threshold using fuzzy logic for cdma system,” in *2011 11th International Symposium on Communications and Information Technologies (ISCIT)*, pp. 142 –147, oct. 2011. 26
- [13] Y. Fang and Y. Zhang, “Call admission control schemes and performance analysis in wireless mobile networks,” *IEEE Transactions on Vehicular Technology*, vol. 51, pp. 371 –382, mar 2002. 26

REFERENCES

- [14] S. Rampal and D. S. Reeves, "Routing and admission control algorithms for multimedia traffic," *Computer Communications*, vol. 18, no. 10, pp. 755 – 768, 1995. 27
- [15] Y. Zhang, J. Luo, and H. Hu, *Wireless Mesh Networking: Architectures, Protocols and Standards*. Auerbach, 2007. 27, 30
- [16] I. F. Akyildiz and X. Wang, "A survey on wireless mesh networks," *IEEE Communications Magazine*, vol. 43, pp. S23–S30, Sept. 2005. 27, 29, 40, 52
- [17] S. Waharte, R. Boutaba, Y. Iraqi, and B. Ishibashi, "Routing protocols in wireless mesh networks: challenges and design considerations," *Multimedia Tools Appl.*, vol. 29, pp. 285–303, June 2006. 28, 41, 44, 45, 47
- [18] D. Johnson, C. Perkins, and J. Arkko, "Mobility support in ipv6," *RFC 3775*, June 2004. 32, 33, 81
- [19] C. Perkins, "Ip mobility support for ipv4," *RFC 3344*, August 2002. 32, 33
- [20] I. Akyildiz, J. Xie, and S. Mohanty, "A survey of mobility management in next-generation all-ip-based wireless systems," *IEEE, Wireless Communications*, vol. 11, pp. 16 – 28, aug. 2004. 33, 35, 80
- [21] A. Campbell, J. Gomez, S. Kim, A. Valko, W. Chieh-Yih, and Z. Turanyi, "Design, implementation, and evaluation of cellular ip," *Personal Communications, IEEE*, vol. 7, pp. 42 –49, 2000. 35, 81
- [22] R. Koodli, "Fast handovers for mobile ipv6," *RFC 4068*, July 2005. 35, 81
- [23] H. Soliman, C. Castelluccia, K. El Malki, and L. Bellier, "Hierarchical mobile ipv6 mobility management (hmipv6)," *RFC 4140*, August 2005. 35, 81

REFERENCES

- [24] P. Reinbold and O. Bonaventure, “Ip micro-mobility protocols,” *IEEE Communications Surveys and Tutorials*, vol. 5,no 1, 2003. 35, 81
- [25] J. Kempf, “Goals for network-based localized mobility management (netlmm),” *RFC 4831*., April 2007. 36
- [26] K.-S. Kong, W. Lee, Y.-H. Han, M.-K. Shin, and H. You, “Mobility management for all-ip mobile networks: mobile ipv6 vs. proxy mobile ipv6,” *Wireless Communications, IEEE*, vol. 15 Issue:2, pp. 36 – 45, April 2008. 37, 81, 84
- [27] C. J. Bernardos and M. Gramaglia, “Network-based localized ip mobility management: Proxy mobile ipv6 and current trends in standardization,” *Journal of Wireless Mobile Networks, Ubiquitous Computing, and Dependable Applications (JoWUA)*, vol. 1, pp. 16 – 35, 2011. 39
- [28] L. Hanzo II and R. Tafazolli, “Admission control schemes for 802.11-based multi-hop mobile ad hoc networks: a survey,” *Communications Surveys Tutorials, IEEE*, vol. 11, pp. 78 –108, quarter 2009. 39, 40
- [29] H. Zhu and I. Chlamtac, “Admission control and bandwidth reservation in multi-hop ad hoc networks,” *Computer Networks*, vol. 50, no. 11, pp. 1653 – 1674, 2006. 40
- [30] Y. Yang and R. Kravets, “Contention-aware admission control for ad hoc networks,” *IEEE Transactions on Mobile Computing*, vol. 4, pp. 363 – 377, July-Aug 2005. 40
- [31] S. Marwaha, J. Indulska, and M. Portmann, “Challenges and recent advances in qos provisioning in wireless mesh networks,” in *8th IEEE International Conference on Computer and Information Technology, 2008. CIT 2008.*, pp. 618 –623, july 2008. 41

REFERENCES

- [32] L. Luo, D. Raychaudhuri, H. Liu, M. Wu, and D. Li, “Joint association, routing and bandwidth allocation for wireless mesh networks,” in *IEEE GLOBECOM 2008*, 2008. 41, 63
- [33] Y. Lin and W. Wong, “An admission control algorithm for multi-hop 802.11 e-based wlans,” *Comput. Commun.*, vol. 31, no. 14, pp. 3510–3520, 2008. 41
- [34] S. Zhang, F. Yu, and V. Leung, “Joint connection admission control and routing in iee 802.16-based mesh networks,” in *IEEE International Conference on Communications, ICC*, May 2008. 42, 65
- [35] S. Lee, N. Girjia, M. Pal, G. Wilfong, and L. Zhang, “Admission control for multihop wireless backhaul networks with qos support,” in *IEEE Wireless Communications and Networking Conference, WCNC 2006*, April 2006. 42, 62, 65
- [36] Y. Yang, J. Wang, and R. Kravets, “Designing routing metrics for mesh networks,” *Proceedings of the IEEE Workshop on Wireless Mesh Networks (WiMesh)*. IEEE Press, 2005. 44, 45
- [37] R. Draves, J. Padhye, and B. Zill, “Routing in multi-radio, multi-hop wireless mesh networks,” in *MobiCom '04: Proceedings of the 10th annual international conference on Mobile computing and networking*, (New York, NY, USA), pp. 114–128, ACM, 2004. 46
- [38] L. Zhao and A. Y. Al-Dubai, “Routing metrics for wireless mesh networks: A survey,” in *Recent Advances in Computer Science and Information Engineering*, Springer Berlin Heidelberg, 2012. 47
- [39] D. P. Bertsekas and R. Gallager, *Data networks (2. ed.)*. Prentice Hall, 1992. 47, 48, 66, 82, 90, 91, 104

REFERENCES

- [40] A. Adya, P. Bahl, J. Padhye, A. Wolman, and L. Zhou, “A multi-radio unification protocol for ieee 802.11 wireless networks,” in *First International Conference on Broadband Networks, 2004. BroadNets 2004.*, pp. 344 – 354, oct. 2004. 47
- [41] S. Keshav, “A control-theoretic approach to flow control,” in *Proceedings of the conference on Communications architecture & protocols*, SIGCOMM ’91, (New York, NY, USA), pp. 3–15, ACM, 1991. 47
- [42] D. S. J. De Couto, D. Aguayo, J. Bicket, and R. Morris, “A high-throughput path metric for multi-hop wireless routing,” *Wirel. Netw.*, vol. 11, pp. 419–434, July 2005. 48
- [43] R. Draves, J. Padhye, and B. Zill, “Routing in multi-radio, multi-hop wireless mesh networks,” in *Proceedings of the 10th annual international conference on Mobile computing and networking*, MobiCom ’04, (New York, NY, USA), pp. 114–128, ACM, 2004. 48
- [44] R. Braden, D. Clark, and S. Shenker, “Integrated services in the internet architecture: an overview,” *RFC 1633*, 1994. 50
- [45] S. Blake, D. Black, M. Carlson, E. Davies, Z. Wang, and W. Weiss, “An architecture for differentiated service,” *RFC 2475*, 1998. 50
- [46] R. Braden, L. Zhang, S. Berson, S. Herzog, and S. Jamin, “Resource reservation protocol (rsvp) – version 1 functional specification,” *RFC 2205*, 1997. 50
- [47] A. E. Taha, H. S. Hassanein, and H. T. Mouftah, “Extensions for internet qos paradigms to mobile ip: a survey,” *Comm. Mag.*, vol. 43, pp. 132–139, May 2005. 50

REFERENCES

- [48] A. Pragad, V. Friderikos, P. Pangalos, and H. Aghvami, "The impact of mobility agent based micro mobility on the capacity of wireless access networks," in *IEEE Global Telecommunications Conference, 2007. GLOBE-COM '07*, pp. 4994–4999, nov. 2007. 51
- [49] A. D. Pragad, V. Friderikos, P. Pangalos, and A. Aghvami, "Qos aware dynamic route optimization for proxy mobile ipv6 networks," *Wireless Communications And Mobile Computing*, vol. 11, p. 508521, 2009. 51, 82, 83, 84, 85, 87, 89, 90, 93, 116, 117
- [50] S. Pack, X. Shen, and J. Mark, "Adaptive route optimization in hierarchical mobile ipv6 networks," *IEEE Transactions on Mobile Computing*, vol. 6, pp. 903–914, aug. 2007. 51, 84
- [51] S. Pack, T. Kwon, and Y. Choi, "Adaptive local route optimization in hierarchical mobile ipv6 networks," in *Wireless Communications and Networking Conference, 2005 IEEE*, vol. 3, pp. 1421–1426 Vol. 3, 2005. 51, 84
- [52] S. Jeong and M.-K. Shin, "Route optimization scheme for proxy mobile ipv6 (pmipv6)," in *Advanced Communication Technology, 2008. ICACT 2008. 10th International Conference on*, vol. 2, pp. 1097–1100, feb. 2008. 51
- [53] M. Baglietto, R. Bolla, F. Davoli, M. Marchese, and M. Mongelli, "A proposal of new price-based call admission control rules for guaranteed performance services multiplexed with best effort traffic," *Comput. Commun.*, vol. 26, pp. 1470–1483, Aug. 2003. 53, 67
- [54] M. Dawande, J. Kalagnanam, P. Keskinocak, R. Ravi, and F. S. Salman, "Approximation algorithms for the multiple knapsack problem with assignment restrictions," *Journal of Combinatorial Optimization*, vol. 4, pp. 171–186, 2000. 53, 61, 70

REFERENCES

- [55] S. Martello and P. Toth, *Knapsack Problems: Algorithms and Computer Implementations*. John Wiley & Sons, 1990. 54, 61, 71
- [56] Q. Fubin and D. Rui, “Simulated annealing for the 0/1 multidimensional knapsack problem,” *Numer. Math. J. Chinese Univ. (English Ser.)*, vol. 16, pp. 320–327, 2007. 54, 72
- [57] S. Nelakuditi, R. Harinath, S. Rayadurgam, and Z.-L. Zhang, “Revenue-based call admission control for wireless cellular networks,” in *1999 IEEE International Conference on Personal Wireless Communication*, pp. 486 – 490, 1999. 56
- [58] O. Yilmaz and I.-R. Chen, “Comparative performance analysis of cac reward optimization algorithms in wireless networks,” in *International Conference on Advanced Information Networking and Applications, AINA '09, 2009.*, pp. 463 –470, may 2009. 56
- [59] I.-R. Chen and N. Verma, “A cost-based admission control algorithm for digital library multimedia systems storing heterogeneous objects,” *Comput. J.*, vol. 46, no. 6, pp. 645–659, 2003. 56
- [60] D. Ayyagari and A. Ephremides, “Admission control with priorities: approaches for multi-rate wireless systems,” in *IEEE 1998 International Conference on Universal Personal Communications, 1998. ICUPC '98.*, vol. 1, pp. 301 –305 vol.1, oct 1998. 56
- [61] S. Nelakuditi, R. Harinath, S. Rayadurgam, and Z.-L. Zhang, “Revenue-based call admission control for wireless cellular networks,” in *1999 IEEE International Conference on Personal Wireless Communication*, 1999. 57
- [62] F. Hou, P.-H. Ho, and X. Shen, “Wlc17-1: Performance analysis of a reservation based connection admission scheme in 802.16 networks,” in *IEEE*

REFERENCES

- Global Telecommunications Conference, 2006. GLOBECOM '06.*, pp. 1 –5, 27 2006-dec. 1 2006. 57
- [63] S.-L. Hew and L. White, “Optimal integrated call admission control and dynamic pricing with handoffs and price-affected arrivals,” in *Asia-Pacific Conference on Communications, 2005*, pp. 396 –400, oct. 2005. 57
- [64] N. Verma and I. Chen, “Admission control algorithms integrated with pricing for revenue optimization with qos guarantees in mobile wireless networks,” in *Tenth International Conference on Parallel and Distributed Systems, 2004. ICPADS 2004.*, pp. 495 – 502, july 2004. 57
- [65] J. Murphy and L. Murphy, “Bandwidth allocation by pricing in atm networks,” in *Proceedings of the IFIP TC6 Second International Conference on Broadband Communications II*, (New York, NY, USA), pp. 333–351, Elsevier Science Inc., 1994. 58
- [66] I. Paschalidis and J. Tsitsiklis, “Congestion-dependent pricing of network services,” *IEEE/ACM Transactions on Networking*, vol. 8, pp. 171 –184, apr 2000. 58
- [67] L. Badia and M. Zorzi, “A framework for call admission control with threshold setup and evaluation of the performance in wcdma systems,” in *Vehicular Technology Conference, 2003. VTC 2003-Spring.*, 2003. 58
- [68] G. Zachariadis and J. Barria, “Dynamic pricing and resource allocation using revenue management for multiservice networks,” *IEEE Transactions on Network and Service Management*, vol. 5, pp. 215 –226, december 2008. 58
- [69] A. Lazaris and P. Koutsakis, “Pricing-based call admission control for wireless mpeg-4 traffic,” *Computer Communications*, vol. 32, no. 17, pp. 1781 – 1789, 2009. 58, 67

REFERENCES

- [70] N. S. Nandiraju, D. S. Nandiraju, and D. P. Agrawal, "Multipath routing in wireless mesh networks," pp. 741–746, oct. 2006. 60, 66
- [71] S. Martello and P. Toth, "Upper bounds and algorithms for hard 0-1 knapsack problems," *Operations Research*, vol. 45, pp. 768–778, September-October 1997. 62, 68
- [72] V. Friderikos and K. Papadaki, "Routing with minimum frame length schedules in wireless mesh networks," in *IEEE Globecom Workshops, 2007*, November 2007. 63
- [73] E. Martins and M. Pascoal, "A new implementation of yens ranking loopless paths algorithm," *Quarterly Journal of the Belgian, French and Italian Operations Research Societies*, vol. 1, pp. 121–133, 2003. 66
- [74] R. Mewanou and S. Pierre, "Dynamic routing algorithms in all-optical networks," in *CCECE*, 2003. 66, 104
- [75] M. Aldebert, M. Ivaldi, and C. Roucolle, "Telecommunications demand and pricing structure: An econometric analysis," in *7th Conf. Telecommunications Systems*, 1999. 67
- [76] O. Yilmaz and I.-R. Chen, "Utilizing call admission control for pricing optimization of multiple service classes in wireless cellular networks," *Comput. Commun.*, vol. 32, no. 2, pp. 317–323, 2009. 67
- [77] M. Randal, G. McMahon, and S. Sugden, "A simulated annealing approach to communication network design," *Combinatorial Optimization*, vol. 6, pp. 55–65, 2002. 72, 73
- [78] S. Jeong, M.-K. Shin, and H.-J. Kim, "Performance analysis of route optimization scheme in proxy mobile ipv6," in *The 12th International Conference*

REFERENCES

- on Advanced Communication Technology (ICACT), 2010*, vol. 1, pp. 180 – 183, feb. 2010. 84
- [79] K. Egevang and P. Francis, “The ip network address translator (nat),” *IETF RFC 1631*, May 1994. 85
- [80] *Managing Peer-To-Peer Traffic With Cisco Service Control Technology*, Cisco White Paper, 2005. 90, 103
- [81] *Cisco Visual Networking Index: Forecast and Methodology, 20112016*, Cisco White Paper, 2012. 90, 103
- [82] A. Pragad, P. Pangalos, V. Friderikos, and A. Aghvami, “Dynamic qos aware route optimization for networks with mobility agents,” in *Consumer Communications and Networking Conference, CCNC 2009*, 2009. 90
- [83] L. Kleinrock and R. Gail, *Queueing Systems: Problems and Solutions*. John Wiley and Sons, 1996. 95, 96

Appendix A

Heuristic and Meta-Heuristic Algorithms

Algorithm 1 Greedy Call Packing (GCP)

```
1: Input:  $w_i, p_i, c_j$  and  $b(v)$ 
2: for  $i = 1$  to  $n$  do
3:    $G = \frac{p_i}{w_i}$ 
4: end for
5: sort the flows so that  $G_1 > G_2 > \dots > G_n$ 
6: set  $L_t, Z_t$  and  $x_{ijk}$  to zero
7: for  $i = 1$  to  $n$  do
8:   find  $j$  for which  $D_{ij} = 1$ 
9:   if  $L_{t_j} + w_i \leq C_j$  then
10:    find which  $sp_k$  has the available capacity
11:    find  $sp^v(jk) = 1$ 
12:    if  $w_i \leq b(v)$  then
13:       $L_{t_j} = L_{t_j} + w_i; Z_t = Z_t + p_j; x_{ijk} = 1;$ 
14:       $b(v) = b(v) - w_i;$ 
15:    end if
16:  end if
17: end for
18: Output  $Z_t$  and  $x_{ijk}$ 
```

Algorithm 2 Simulated Annealing (SA)

```

1: select the initial solution :  $I_o = \{ijk \mid x_{ijk} = 0\}$  and  $I_1 = \{ijk \mid x_j = 1\}$  for
    $\{j = 1, \dots, M; i = 1, \dots, n; k = 1, \dots, K\}$ , initial temperature  $t_o$ , number of
   iterations per temperature  $tr$ , cooling schedule  $D$  and final temperature  $t_f$ .
2:  $t = t_o$ 
3: repeat
4:   repeat
5:     set iteration counter  $no = 0$ ;
6:     choose  $h \in I_o$  randomly
7:     find  $D_{hj} = 1$ 
8:     if  $L_{t_j} + w_h \leq c_j$  then
9:       find  $sp^v(jk) = 1$ 
10:      if  $w_h \leq b(v)$  then
11:         $I_o = I_o - h; I_1 = I_1 \cup h; b(v) = b(v) - w_h$ ;
12:         $no = tr$ ; goto cool
13:      end if
14:    end if
15:    choose  $k \in I_1$  that is admitted to  $MAP_j$ 
16:    find  $D_{kj} = 1$ 
17:    if  $L_{t_j} + (w_h - w_k) \leq c_j$  then
18:      find  $k$  and  $sp^v(jk) = 1$ 
19:      if  $w_h - w_k \leq b(v)$  then
20:         $\delta = p_h - p_k$ ;
21:        if  $\delta > 0$  then
22:           $I_o = I_o - h \cup k; I_1 = I_1 \cup h - k; Z_t = Z_t + \delta$ ;
23:           $no = tr$ ; goto cool
24:        else
25:           $\delta f = Z_t + \delta$ ;
26:          if randomly distributed  $y \text{ in}(0, 1) < \exp(\frac{-\delta f}{t})$  then
27:             $I_o = I_o - h \cup k; I_1 = I_1 \cup h - k; Z_t = Z_t + \delta$ ;
28:             $no = tr$ ; goto cool
29:          end if
30:        end if
31:      end if
32:    else
33:      if randomly distributed  $y \text{ in}(0, 1) < \exp(\frac{-p_k}{t})$  then
34:         $I_o = I_o \cup k; I_1 = I_1 - k; Z_t = Z_t - p_k$ ;
35:         $no = tr$ ; goto cool
36:      end if
37:    end if
38:     $no = no + 1$ ;
39:  until  $no = tr$ ;
40:  cool :  $t = t \times D$ ;
41: until  $t < t_f$ 

```
